

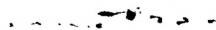
Fr m: Mertz, Prema  
Sent: Thursday, January 24, 2002 10:24 AM  
T : STIC-Biotech/ChemLib  
Subject: 09/521,195

Please search SEQ ID NO:1 and 3 with protein databases.

Thanks  
Prema Mertz, Ph.D.  
Art Unit 1646  
Mailbox 10C-01  
Crystal Mall 1, Room 10E-01  
United States Patent & Trademark Office  
# (703) 308-4229

Edward Hart  
Technical Info Specialist  
STIC / Biotech  
CM1 12C14 Tel: 305-9203

TYPE OF SEARCH:		VENDOR/COST(where applic.)
Searcher: _____	NA Sequences: _____	STN: _____
Phone: _____	AA Sequences: <u>2</u> _____	DIALOG: _____
Location: _____	Structures: _____	Questel/Orbit: _____
Date Picked Up: <u>1/25/02</u>	Bibliographic: _____	DRLink: _____
Date Completed: <u>1/25/02</u>	Litigation: _____	Lexis/Nexis: _____
Searcher Prep/Review: _____	Full text: _____	Sequence Sys.: <u>Q2</u> _____
Clerical: _____	Patent Family: _____	WWW/Internet: _____
Online time: _____	Other: _____	Other (specify): _____



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

protein - protein search, using sw model

on January 25, 2002, 10:08:19 ; Search time 43.11 Seconds

(without alignments)  
946.750 Million cell updates/sec

US-09-521-195-1  
score: 2845  
e 1 WRDYDEVIAFGWGPQRL.....KKTRDSMETENPKVLITAF 551

, table: BLOSUM62

ied 522463 segs, 74073290 residues  
number of hits satisfying chosen parameters: 522463

um DB seq length: 0

-processing: Minimum Match 0%

Listing first 45 summaries

abase : A\_Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than the observed score of the best hit. The results are sorted by the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2845	100.0	551	20 AAY01649	A protein with cat
2	2470	86.8	553	20 AAY01651	A protein with cat
3	2470	86.8	553	21 AAB20579	Mouse OCTN1 amino
4	2218	78.0	557	20 AAY01650	A protein with cat
5	2218	78.0	557	21 AAY83929	Human carnitine tr
6	2164	76.1	557	20 AAY01651	Protein with cat
7	2164	76.1	557	20 AAB20580	Mouse OCTN2
8	2164	76.1	557	21 AAY83930	Mouse carnitine tr
9	2034	5	564	21 AAB20578	Mouse OCTN3 protei
10	779	5	560	21 AAB43038	Human OREF2802
11	779	5	584	22 AAM00930	Human bone marrow

12	698	24.5	540	22 AAB49401	Murine organic ani
13	695.5	24.1	554	19 AAW64538	Human liver cell
14	692	24.0	535	21 AAY51249	Rat liver anion tr
15	681.5	23.4	556	17 AAB77676	Rat OCT-1 protein
16	664.5	23.3	546	22 AAB47272	HOAT2A. Homo sapi
17	663.5	23.3	538	22 AAB47273	HOAT2B. Homo sapi
18	652.5	22.9	545	21 AAY83929	Human organic an
19	652.5	22.9	545	21 AAY83929	Human organic an
20	650	22.8	551	21 AAB08824	A human organic an
21	647	22.7	542	22 AAB47274	HOAT3. Homo sapi
22	637	22.4	536	21 AAY92903	Rat cerebral organ
23	636	22.4	561	18 AAW44196	Human osteoclast
24	630	22.1	537	18 AAW44195	Mouse osteoclast
25	624	21.8	578	22 AAB06571	Human protein havi
26	621	21.8	578	22 AAB06571	Human protein havi
27	614.5	21.5	483	22 AAM00982	Mouse bone marrow
28	606.5	21.3	545	22 AAB36553	Mouse organic anio
29	597.5	21.0	551	20 AAW88488	Human organic anio
30	586	20.6	550	21 AAY44278	HOAT1. Homo sapi
31	583.5	20.5	520	22 AAB84749	Human organic anio
32	583.5	20.5	520	22 AAB84749	Human organic anio
33	564	19.8	550	22 AAE06612	Human protein havi
34	564	19.8	550	22 AAB69091	Human organic anio
35	471.5	16.6	541	22 AAB47276	HOAT5. Homo sapi
36	430.5	15.1	553	22 AAE04897	Human transporter
37	417	14.7	480	21 AAG28044	Arabidopsis thalia
38	417	14.7	480	21 AAG28044	Arabidopsis thalia
39	412.5	14.5	521	21 AAG30895	Arabidopsis thalia
40	402	14.1	554	22 AAB47275	HOAT4. Homo sapi
41	387.5	13.6	397	21 AAG30886	Arabidopsis thalia
42	382.5	13.4	515	21 AAG42522	Arabidopsis thalia
43	341	12.0	360	21 AAG30887	Arabidopsis thalia
44	345.5	11.4	378	21 AAG42523	Arabidopsis thalia
45	312.5	11.0	359	21 AAG42524	Arabidopsis thalia

## ALIGNMENTS

RESULT 1  
AAY01649  
ID AAY01649 standard; Protein: 551 AA.  
AC AAY01649:  
DT 23-JUN-1999 (first entry)  
DE A protein with cation transporting activity.  
KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver,  
heart disease; cancer; anti-tumour drug; anticancer drug.  
XX Homo sapiens.  
XX WO9913072-A1.  
XX PD 18-MAR-1999.  
XX PF 07-SEP-1998; 98WO-JP04009.  
XX PR 20-MAY-1999; 98JP-015660.  
XX PR 08-SEP-1997; 97JP-026092.  
XX PA (CHUGAI) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX PI Nezu J, Oku A;  
XX WP1: 1999-215062/18.  
XX WP2: 1999-215062/18.  
XX WP3: 1999-215062/18.  
XX WP4: 1999-215062/18.  
XX WP5: 1999-215062/18.  
XX WP6: 1999-215062/18.  
XX WP7: 1999-215062/18.  
XX WP8: 1999-215062/18.  
XX WP9: 1999-215062/18.  
XX WP10: 1999-215062/18.  
XX WP11: 1999-215062/18.  
XX WP12: 1999-215062/18.  
XX WP13: 1999-215062/18.  
XX WP14: 1999-215062/18.  
XX WP15: 1999-215062/18.  
XX WP16: 1999-215062/18.  
XX WP17: 1999-215062/18.  
XX WP18: 1999-215062/18.  
XX WP19: 1999-215062/18.  
XX WP20: 1999-215062/18.  
XX WP21: 1999-215062/18.  
XX WP22: 1999-215062/18.  
XX WP23: 1999-215062/18.  
XX WP24: 1999-215062/18.  
XX WP25: 1999-215062/18.  
XX WP26: 1999-215062/18.  
XX WP27: 1999-215062/18.  
XX WP28: 1999-215062/18.  
XX WP29: 1999-215062/18.  
XX WP30: 1999-215062/18.  
XX WP31: 1999-215062/18.  
XX WP32: 1999-215062/18.  
XX WP33: 1999-215062/18.  
XX WP34: 1999-215062/18.  
XX WP35: 1999-215062/18.  
XX WP36: 1999-215062/18.  
XX WP37: 1999-215062/18.  
XX WP38: 1999-215062/18.  
XX WP39: 1999-215062/18.  
XX WP40: 1999-215062/18.  
XX WP41: 1999-215062/18.  
XX WP42: 1999-215062/18.  
XX WP43: 1999-215062/18.  
XX WP44: 1999-215062/18.  
XX WP45: 1999-215062/18.  
XX WP46: 1999-215062/18.  
XX WP47: 1999-215062/18.  
XX WP48: 1999-215062/18.  
XX WP49: 1999-215062/18.  
XX WP50: 1999-215062/18.  
XX WP51: 1999-215062/18.  
XX WP52: 1999-215062/18.  
XX WP53: 1999-215062/18.  
XX WP54: 1999-215062/18.  
XX WP55: 1999-215062/18.  
XX WP56: 1999-215062/18.  
XX WP57: 1999-215062/18.  
XX WP58: 1999-215062/18.  
XX WP59: 1999-215062/18.  
XX WP60: 1999-215062/18.  
XX WP61: 1999-215062/18.  
XX WP62: 1999-215062/18.  
XX WP63: 1999-215062/18.  
XX WP64: 1999-215062/18.  
XX WP65: 1999-215062/18.  
XX WP66: 1999-215062/18.  
XX WP67: 1999-215062/18.  
XX WP68: 1999-215062/18.  
XX WP69: 1999-215062/18.  
XX WP70: 1999-215062/18.  
XX WP71: 1999-215062/18.  
XX WP72: 1999-215062/18.  
XX WP73: 1999-215062/18.  
XX WP74: 1999-215062/18.  
XX WP75: 1999-215062/18.  
XX WP76: 1999-215062/18.  
XX WP77: 1999-215062/18.  
XX WP78: 1999-215062/18.  
XX WP79: 1999-215062/18.  
XX WP80: 1999-215062/18.  
XX WP81: 1999-215062/18.  
XX WP82: 1999-215062/18.  
XX WP83: 1999-215062/18.  
XX WP84: 1999-215062/18.  
XX WP85: 1999-215062/18.  
XX WP86: 1999-215062/18.  
XX WP87: 1999-215062/18.  
XX WP88: 1999-215062/18.  
XX WP89: 1999-215062/18.  
XX WP90: 1999-215062/18.  
XX WP91: 1999-215062/18.  
XX WP92: 1999-215062/18.  
XX WP93: 1999-215062/18.  
XX WP94: 1999-215062/18.  
XX WP95: 1999-215062/18.  
XX WP96: 1999-215062/18.  
XX WP97: 1999-215062/18.  
XX WP98: 1999-215062/18.  
XX WP99: 1999-215062/18.  
XX WP100: 1999-215062/18.  
XX WP101: 1999-215062/18.  
XX WP102: 1999-215062/18.  
XX WP103: 1999-215062/18.  
XX WP104: 1999-215062/18.  
XX WP105: 1999-215062/18.  
XX WP106: 1999-215062/18.  
XX WP107: 1999-215062/18.  
XX WP108: 1999-215062/18.  
XX WP109: 1999-215062/18.  
XX WP110: 1999-215062/18.  
XX WP111: 1999-215062/18.  
XX WP112: 1999-215062/18.  
XX WP113: 1999-215062/18.  
XX WP114: 1999-215062/18.  
XX WP115: 1999-215062/18.  
XX WP116: 1999-215062/18.  
XX WP117: 1999-215062/18.  
XX WP118: 1999-215062/18.  
XX WP119: 1999-215062/18.  
XX WP120: 1999-215062/18.  
XX WP121: 1999-215062/18.  
XX WP122: 1999-215062/18.  
XX WP123: 1999-215062/18.  
XX WP124: 1999-215062/18.  
XX WP125: 1999-215062/18.  
XX WP126: 1999-215062/18.  
XX WP127: 1999-215062/18.  
XX WP128: 1999-215062/18.  
XX WP129: 1999-215062/18.  
XX WP130: 1999-215062/18.  
XX WP131: 1999-215062/18.  
XX WP132: 1999-215062/18.  
XX WP133: 1999-215062/18.  
XX WP134: 1999-215062/18.  
XX WP135: 1999-215062/18.  
XX WP136: 1999-215062/18.  
XX WP137: 1999-215062/18.  
XX WP138: 1999-215062/18.  
XX WP139: 1999-215062/18.  
XX WP140: 1999-215062/18.  
XX WP141: 1999-215062/18.  
XX WP142: 1999-215062/18.  
XX WP143: 1999-215062/18.  
XX WP144: 1999-215062/18.  
XX WP145: 1999-215062/18.  
XX WP146: 1999-215062/18.  
XX WP147: 1999-215062/18.  
XX WP148: 1999-215062/18.  
XX WP149: 1999-215062/18.  
XX WP150: 1999-215062/18.  
XX WP151: 1999-215062/18.  
XX WP152: 1999-215062/18.  
XX WP153: 1999-215062/18.  
XX WP154: 1999-215062/18.  
XX WP155: 1999-215062/18.  
XX WP156: 1999-215062/18.  
XX WP157: 1999-215062/18.  
XX WP158: 1999-215062/18.  
XX WP159: 1999-215062/18.  
XX WP160: 1999-215062/18.  
XX WP161: 1999-215062/18.  
XX WP162: 1999-215062/18.  
XX WP163: 1999-215062/18.  
XX WP164: 1999-215062/18.  
XX WP165: 1999-215062/18.  
XX WP166: 1999-215062/18.  
XX WP167: 1999-215062/18.  
XX WP168: 1999-215062/18.  
XX WP169: 1999-215062/18.  
XX WP170: 1999-215062/18.  
XX WP171: 1999-215062/18.  
XX WP172: 1999-215062/18.  
XX WP173: 1999-215062/18.  
XX WP174: 1999-215062/18.  
XX WP175: 1999-215062/18.  
XX WP176: 1999-215062/18.  
XX WP177: 1999-215062/18.  
XX WP178: 1999-215062/18.  
XX WP179: 1999-215062/18.  
XX WP180: 1999-215062/18.  
XX WP181: 1999-215062/18.  
XX WP182: 1999-215062/18.  
XX WP183: 1999-215062/18.  
XX WP184: 1999-215062/18.  
XX WP185: 1999-215062/18.  
XX WP186: 1999-215062/18.  
XX WP187: 1999-215062/18.  
XX WP188: 1999-215062/18.  
XX WP189: 1999-215062/18.  
XX WP190: 1999-215062/18.  
XX WP191: 1999-215062/18.  
XX WP192: 1999-215062/18.  
XX WP193: 1999-215062/18.  
XX WP194: 1999-215062/18.  
XX WP195: 1999-215062/18.  
XX WP196: 1999-215062/18.  
XX WP197: 1999-215062/18.  
XX WP198: 1999-215062/18.  
XX WP199: 1999-215062/18.  
XX WP200: 1999-215062/18.  
XX WP201: 1999-215062/18.  
XX WP202: 1999-215062/18.  
XX WP203: 1999-215062/18.  
XX WP204: 1999-215062/18.  
XX WP205: 1999-215062/18.  
XX WP206: 1999-215062/18.  
XX WP207: 1999-215062/18.  
XX WP208: 1999-215062/18.  
XX WP209: 1999-215062/18.  
XX WP210: 1999-215062/18.  
XX WP211: 1999-215062/18.  
XX WP212: 1999-215062/18.  
XX WP213: 1999-215062/18.  
XX WP214: 1999-215062/18.  
XX WP215: 1999-215062/18.  
XX WP216: 1999-215062/18.  
XX WP217: 1999-215062/18.  
XX WP218: 1999-215062/18.  
XX WP219: 1999-215062/18.  
XX WP220: 1999-215062/18.  
XX WP221: 1999-215062/18.  
XX WP222: 1999-215062/18.  
XX WP223: 1999-215062/18.  
XX WP224: 1999-215062/18.  
XX WP225: 1999-215062/18.  
XX WP226: 1999-215062/18.  
XX WP227: 1999-215062/18.  
XX WP228: 1999-215062/18.  
XX WP229: 1999-215062/18.  
XX WP230: 1999-215062/18.  
XX WP231: 1999-215062/18.  
XX WP232: 1999-215062/18.  
XX WP233: 1999-215062/18.  
XX WP234: 1999-215062/18.  
XX WP235: 1999-215062/18.  
XX WP236: 1999-215062/18.  
XX WP237: 1999-215062/18.  
XX WP238: 1999-215062/18.  
XX WP239: 1999-215062/18.  
XX WP240: 1999-215062/18.  
XX WP241: 1999-215062/18.  
XX WP242: 1999-215062/18.  
XX WP243: 1999-215062/18.  
XX WP244: 1999-215062/18.  
XX WP245: 1999-215062/18.  
XX WP246: 1999-215062/18.  
XX WP247: 1999-215062/18.  
XX WP248: 1999-215062/18.  
XX WP249: 1999-215062/18.  
XX WP250: 1999-215062/18.  
XX WP251: 1999-215062/18.  
XX WP252: 1999-215062/18.  
XX WP253: 1999-215062/18.  
XX WP254: 1999-215062/18.  
XX WP255: 1999-215062/18.  
XX WP256: 1999-215062/18.  
XX WP257: 1999-215062/18.  
XX WP258: 1999-215062/18.  
XX WP259: 1999-215062/18.  
XX WP260: 1999-215062/18.  
XX WP261: 1999-215062/18.  
XX WP262: 1999-215062/18.  
XX WP263: 1999-215062/18.  
XX WP264: 1999-215062/18.  
XX WP265: 1999-215062/18.  
XX WP266: 1999-215062/18.  
XX WP267: 1999-215062/18.  
XX WP268: 1999-215062/18.  
XX WP269: 1999-215062/18.  
XX WP270: 1999-215062/18.  
XX WP271: 1999-215062/18.  
XX WP272: 1999-215062/18.  
XX WP273: 1999-215062/18.  
XX WP274: 1999-215062/18.  
XX WP275: 1999-215062/18.  
XX WP276: 1999-215062/18.  
XX WP277: 1999-215062/18.  
XX WP278: 1999-215062/18.  
XX WP279: 1999-215062/18.  
XX WP280: 1999-215062/18.  
XX WP281: 1999-215062/18.  
XX WP282: 1999-215062/18.  
XX WP283: 1999-215062/18.  
XX WP284: 1999-215062/18.  
XX WP285: 1999-215062/18.  
XX WP286: 1999-215062/18.  
XX WP287: 1999-215062/18.  
XX WP288: 1999-215062/18.  
XX WP289: 1999-215062/18.  
XX WP290: 1999-215062/18.  
XX WP291: 1999-215062/18.  
XX WP292: 1999-215062/18.  
XX WP293: 1999-215062/18.  
XX WP294: 1999-215062/18.  
XX WP295: 1999-215062/18.  
XX WP296: 1999-215062/18.  
XX WP297: 1999-215062/18.  
XX WP298: 1999-215062/18.  
XX WP299: 1999-215062/18.  
XX WP300: 1999-215062/18.  
XX WP301: 1999-215062/18.  
XX WP302: 1999-215062/18.  
XX WP303: 1999-215062/18.  
XX WP304: 1999-215062/18.  
XX WP305: 1999-215062/18.  
XX WP306: 1999-215062/18.  
XX WP307: 1999-215062/18.  
XX WP308: 1999-215062/18.  
XX WP309: 1999-215062/18.  
XX WP310: 1999-215062/18.  
XX WP311: 1999-215062/18.  
XX WP312: 1999-215062/18.  
XX WP313: 1999-215062/18.  
XX WP314: 1999-215062/18.  
XX WP315: 1999-215062/18.  
XX WP316: 1999-215062/18.  
XX WP317: 1999-215062/18.  
XX WP318: 1999-215062/18.  
XX WP319: 1999-215062/18.  
XX WP320: 1999-215062/18.  
XX WP321: 1999-215062/18.  
XX WP322: 1999-215062/18.  
XX WP323: 1999-

XX Claim 1, Page 41-45; 97pp; Japanese.  
 XX The present sequence represents a protein with cation transporting  
 CC activity. The genes are significantly homologous with organic cation  
 CC transporters OCT1 and OCT2. The genes may be used in drug development,  
 CC particularly in the treatment of diseases due to abnormality of the  
 CC organic cation transporter functions e.g. fatty liver, heart diseases  
 CC and cancers, by controlling such as by inhibition or activation.  
 CC Administration of anti-tumour and anticancer drugs in combination with  
 CC a transporter protein inhibiting agent allows the agents to penetrate  
 CC into the diseased cells to enhance the drug action.  
 XX Sequence 551 AA;  
 SQ

Query Match 100.0%; Score 2845; DB 20; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-281; Indels 0; Gaps 0;  
 Matches 551; Conservative 0; Mismatches 0;  
 QY 1 MRDYDEVIAFLGEGWPTQRLIFLLSASIPNGNGMSVVFAGTPEHRCRVPDAANLSS 60  
 Db 1 mrdydeviaflgwgppqrlifllsasilpngngmsvvlagtpehrcrvpdaanlss 60  
 QY 61 AWRNNSVPLRLDGRREVPHSCRYLATANFSALGLEPGRDVLGQESCLDGEWFS 120  
 Db 61 awrnnsvplrlldgrrevphscrylatanfsalglepgrdvlgqesclgdwefs 120  
 QY 121 ODVYLSVVTWNLVCEDNKVPYLTSLFFVGLGVSFVSGQLSDRFGRKNVLFATMAVQ 180  
 Db 121 odvylsvvtwnlvceadnkvpyltstlffvglgvsfsvsgqlsdrfgrknvlfatmavq 180  
 QY 181 TGFSTLQFISGEMFTVLEIVGNGOISNVVAFILGTLGKSVRIIPSLGVCCTEFA 240  
 Db 181 tgfstlqfisisemftvleivgngoisnvvafilgtlkgsvrliipslgvccteffa 240  
 Y 241 VGYMLLPFAFIRDMWMLLLATVPGVLCVPLWFIPESPRWLISQRFEADIIQKA 300  
 241 vgymlplfayfirdwmlllatvpvclcvplwfiipesprwlisqrfeadiiqlka 300  
 301 AKNNNTAVPAVIFDSVE--ELNPLKOKAFILDLPRTNIAIMTSLMLMLTSVGYFALS 360  
 301 aknnntavpavifdsveel-nplkokaflldlprtniaimtslmlmltsvgyfals 360  
 361 LDAPNLHGDAYLNCPLSALIEIPAYITAMILLRTLPRTYIAAVLFWGGVLLFIQLVPV 420  
 361 ldepnlhgdylncplsaliiepaitamillrtlprtyiaavlfgggvllfiqlvpv 420  
 421 DYVSLTGLVMGKGCITSFMSLVYFAELYPTLVNMAVGVTASTASVGSIIAPYFY 480  
 421 dyvsltglmvgkgcitsfmslvfyfaelyptlvnmavgvtastasvgsiiapyfy 480  
 481 LGAYNMLPYVNGSLAVLIGIFTLPPELGMVLPETLEQMKVWFRSGKTRDSMET 540  
 481 lgaynmlpyvngslavlifgtlppelgmvlpetleqmkvwfrsgktrdsmet 540  
 541 EENPKVLITAF 551  
 541 eenpkvlitaf 551

2  
 AAY01651 standard; Protein; 553 AA.  
 AAY01651,  
 23-JUN-1999 (first entry)  
 A protein with cation transporting activity.  
 Organic cation transporter; OCT1; OCT2; drug development; fatty liver;  
 heart disease; cancer; anti-tumour drug; anticancer drug.  
 4

XX Mus musculus.  
 XX WO9913072-AL.  
 XX 18-MAR-1999.  
 XX 07-SEP-1998; 98WO-JP04009.  
 XX 20-MAY-1998; 98JP-0156660.  
 XX 08-SEP-1997; 97JP-0260972.  
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX Nezu J, Oku A;  
 XX WPI: 1999-215062/18.  
 XX N-PSDB: AAX26898.  
 PT Genes homologous with organic cation transporters OCT1 and OCT2,  
 PT useful in design of new drugs for treatment of diseases due to  
 PT abnormality of the transporter functions  
 XX Claim 1; Page 63-67; 97pp; Japanese.  
 CC The present sequence represents a protein with cation transporting  
 CC activity. The genes are significantly homologous with organic cation  
 CC transporters OCT1 and OCT2. The genes may be used in drug development,  
 CC particularly in the treatment of diseases due to abnormality of the  
 CC organic cation transporter functions e.g. fatty liver, heart diseases  
 CC and cancers, by controlling such as by inhibition or activation.  
 CC Administration of anti-tumour and anticancer drugs in combination with  
 CC a transporter protein inhibiting agent allows the agents to penetrate  
 CC into the diseased cells to enhance the drug action.  
 XX Sequence 553 AA;  
 SQ

Query Match 86.8%; Score 2470; DB 20; Length 553;  
 Best Local Similarity 84.6%; Pred. No. 2.9e-243;  
 Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;  
 QY 1 MRDYDEVIAFLGEGWPTQRLIFLLSASIPNGNGMSVVFAGTPEHRCRVPDAANLSS 60  
 Db 1 mrdydeviaflgwgppqrlifllsasilpngngmsvvlagtpehrcrvpdtvniss 60  
 QY 61 AWRNNSVPLRLDGRREVPHSCRYLATANFSALGLEPGRDVLGQESCLDGEWFS 120  
 Db 61 awrnnsvplrlldgrrevphscrylatanfsalglepgrdvlgqesclgdwef 120  
 QY 121 ODVYLSVVTWNLVCEDNKVPYLTSLFFVGLGVSFVSGQLSDRFGRKNVLFATMAVQ 180  
 Db 121 odvylsvvtwnlvceadnkvpyltstlffvglgvsfsvsgqlsdrfgrknvlfatmavq 180  
 QY 181 TGFSTLQFISGEMFTVLEIVGNGOISNVVAFILGTLGKSVRIIPSLGVCCTEFA 240  
 Db 181 tgfstlqfisisemftvleivgngoisnvvafilgtlkgsvrliipslgvccteffa 240  
 QY 241 VGYMLLPFAFIRDMWMLLLATVPGVLCVPLWFIPESPRWLISQRFEADIIQKA 300  
 Db 241 vgymlplfayfirdwmlllatvpvclcvplwfiipesprwlisqrfeadiiqlka 300  
 QY 301 AKNNNTAVPAVIFDSVE--ELNPLKOKAFILDLPRTNIAIMTSLMLMLTSVGYFALS 358  
 Db 301 aknnntavpavifdsve-el-nplkokaflldlprtniaimtslmlmltsvgyfa 360  
 QY 359 LSLDAPNLHGDAYLNCPLSALIEIPAYITAMILLRTLPRTYIAAVLFWGGVLLFIQLV 418  
 Db 359 lsldepnlhgdylncplsaliiepaitamillrtlprtyiaavlfgggvllfiqlv 420  
 QY 419 PDVYFSLTGLVMGKGCITSFMSLVYFAELYPTLVNMAVGVTASTASVGSIIAPYFY 478  
 Db 419 pdvyfsltglmvgkgcitsfmslvfyfaelyptlvnmavgvtastasvgsiiapyf 480  
 Db 421 PEDYNFVSLGVMGKGCITSFMSLVYFAELYPTLVNMAVGVTASTASVGSIIAPYFY 480



QY 479 VYLGAYNRMPLPYVMGSLVTLGIFTLPPPSIGMTLPETLEOMQKVKWFRSGKKTDRSM 538  
 DB 481 VYLGAYNRMPLPYVMGSLVTLGIFTLPPPSIGMTLPETLEOMQKVKWFRSGKKTDRSM 540

QY 539 ETEENPKVLITAF 551  
 DB 541 dreepkvlitaf 553

## RESULT 3

AAB20579 AAB20579 standard; Protein; 553 AA.

XX AAB20579;

DT 11-DEC-2000 (first entry)

XX Mouse OCTN1 amino acid sequence.

DE Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;  
 identification; regulator; carnitine transport.

XX Mus musculus.

PN WO200046368-A1.

XX 10-AUG-2000.

XX 04-FEB-2000; 2000WO-JP00619.

XX 05-FEB-1999; 99JP-0028406.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Nezu J, Ose A;

XX WPI; 2000-586982/55.

XX Organic cation transporter gene OCTN3 expressed in testis for  
 identification of regulators of carnitine transport for use as drugs

XX Example 2; Fig 2; 58pp; Japanese.

XX The present invention describes a mouse organic cation transporter  
 protein (OCTN3). Also described are: (1) a method for screening  
 compounds for their ability to regulate the transport of an organic  
 cation into the cell, by generating a cell expressing OCTN3 at the  
 cell membrane, contacting with the compound and organic cation, and  
 observing the degree of transport of the organic cation; and (2) a  
 method for screening compounds for their ability to be transported into  
 the cell by OCTN3, by generating a cell expressing OCTN3 at the cell  
 membrane, contacting with the compound and observing the degree of  
 transport of the compound. OCTN3 can be used for the identification of  
 regulators of the transport of organic cations (especially carnitine)  
 into cells by OCTN3, for use as drugs. The present sequence represents  
 the mouse OCTN1 amino acid sequence, which is used in an example from  
 the present invention.

XX Sequence 553 AA;

Query Match 86.8%; Score 2470; DB 21; Length 553;  
 Best Local Similarity 84.8%; Pred. No. 2.9e-243;  
 Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

QY 1 MRDYDEVIAFLGEGPQRLLFFLLSASIPNGFMGSMVVFVLAGTPEHRCRVPDANLSS 60  
 DB 1 mrdydeviaflgwegpqrllffllsasiipngfmgsmvfvlagtpehrcrlvpdtnlss 60

QY 61 AWRNNVPLRLRDCREVPHSCRYRLATIANFSALGCEGRDVLGOLESCLDGWEFS 120  
 DB 61 swrnshp1etckdgrvpggscrryrlatiansamglepggdvldieqescldgwey 120

QY 121 QDVYLTSTVTEWNLVCEDENNKVPJTTLSPFVGLGFSVSGQLSDREGKKNVLPATMAVQ 180  
 DB 121 kdlfistlvtewnlvceeddwnktptltslfvfgvicsfvgqlsdrfgkknvlfatmavq 180  
 QY 181 TGFSFLQIFSSWEMFTVLFVIVMGGOISNYVAVFILTGLTGKSVRIIFSTLGVCTFFA 240  
 DB 181 tgfsfvqlfstnwmftvlfvfaivmgqisnyvavfilitglteilsksvriifstlgvctffa 240  
 QY 241 VCYMLPLPFAVFIQDWRLMLLALTVCVLCVPLWMTIPESRWLISORFREAEDIQKA 300  
 DB 241 igymvlpifayfirdwrmllaltvcplwmtipesrwlsqrfaeaediqka 300  
 QY 301 AKMNTAVPAVIFDSVE--ELNPLKQOKAFILDLFTRNIAIMTINSLMLWMTSVGYFA 358  
 DB 301 akmsnivapagifdplelqelnsikqkvilldlfrtniatitvmavmlwmtsvgyfa 360  
 QY 359 LSLDAPNLHGDAYLNCPLSALIEIPAVITAWLLRLTPRYITAAVLFWGGVLLFIQIV 418  
 DB 361 lslnvpnlhgdvynclflsglievpayftawllrltpryiaagvltwggvllliqv 420  
 QY 419 PVDYFSLSIGLVMLGKFGITSAFSLMYVFTAELYPTLVNRNAVGTSTASRVGSIAPYF 478  
 DB 421 pedyfsvsiglvmlgkfgitsafsmlyvftaelypptlvnrnavgitsmasrvgsiapyf 480  
 QY 479 VYLGAYNRMPLPYVMGSLVTLGIFTLPPPSIGMTLPETLEOMQKVKWFRSGKKTDRSM 538  
 DB 481 vylgaynrmplpyilmsgslvtlglitffpesigvtlpenleqmkgvrgircgkktvs 540  
 QY 539 ETEENPKVLITAF 551  
 DB 541 dreepkvlitaf 553

## RESULT 4

AAY01650 AAY01650 standard; Protein; 557 AA.

XX AC AAY01650;

XX DT 23-JUN-1999 (first entry)

XX A protein with cation transporting activity.

XX Organic cation transporter; OCT1; OCT2; drug development; fatty liver;  
 heart disease; cancer; anti-tumour drug; anticancer drug.

XX Homo sapiens.

XX WO9913072-A1.

XX 18-MAR-1999.

XX 07-SEP-1998; 98WO-JP04009.

XX 20-MAY-1998; 98JP-0156660.

XX 08-SEP-1997; 97JP-0260972.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Nezu J, Ose A;

XX WPI; 1999-215062/18.

XX N-PSDB; AAX26880.

XX Genes homologous with organic cation transporters OCT1 and OCT2,  
 useful in design of new drugs for treatment of diseases due to  
 abnormality of the transporter functions

XX Claim 1; Page 51-55; 97pp; Japanese.

XX The present sequence represents a protein with cation transporting

XX activity. The genes are significantly homologous with organic cation

CC transporters OCT1 and OCT2. The genes may used in drug development,  
 CC particularly in the treatment of diseases due to abnormality of the  
 CC organic cation transporter functions e.g. fatty liver, heart diseases  
 CC and cancers, by controlling such as by inhibition or activation.  
 CC Administration of anti-tumour and anticancer drugs in combination with  
 CC a transporter protein inhibiting agent allows the agents to penetrate  
 CC into the diseased cells to enhance the drug action.  
 XX  
 XX Sequence 557 AA;  
 SQ

Query Match 78.0%; Score 2218; DB 20; Length 557;  
 Best Local Similarity 75.9%; Pred. No. 1.5e-217;  
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

QY 1 MRDYDEVIAFLGEMGPFORLFFLLSASIPNGFNGMSVFLAGTPEHRCRVPDANLSS 60  
 DB 1 mrdydeviaflgawpffgrllffllsasiipngftgssvflattpehrcrvpdaanlss 60  
 QY 61 AWNNNSVPLRLRGREVPSCSRVRLATIANFSALEPGRDVDLQLEQESCLDGEWFS 120  
 DB 61 awnhtvplrlrdgrevphscrryrlatiansaalgldpgrdvdldglegescldgweifs 120  
 QY 121 QDYLSTVTVTEWNLNCEWNVPLTTSILFFVGLGSEFVSQSLSDRFGKKNVLEATMAVQ 180  
 DB 121 qdylstivtewnlvnceddwkapitlsilffvgvllgsfsgqslsdrgfrknvlftmgnmq 180  
 QY 181 TGFSLQIFSIWENFTVLFVIVGMOISNVYVAFILGTEILGKSVRIIFSTLGCTFFA 240  
 DB 181 tgsfllqifsknfemfvvlvlgmgqisnyvaafvlgteilgksvriifstlgvcifya 240  
 QY 241 VGMULLPLFAYEIRDRWMLLALAVPGVLCVPLWVWTFESPRWLSIORFREAEITQKA 300  
 DB 241 fgmvlplfayfirdwrmllalavpgvllcvplwvwtvfpesprwlsiorsrfreaeditqka 300  
 QY 301 AKNNNTAVPAVIFD--SVEELNPLKQKAFILDLFRTNIAIMTINSLMLMTSVGYFA 358  
 DB 301 akangivvpstifdpseqlqdlsskkqgshnldlrltwnlrmvtimsimlwtisvgyfg 360  
 QY 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAAVFLWGGVGLFIQIV 418  
 DB 361 lsldpnlhgdifvncflsalamvepayvawlllqylprysmatalfggavllfmqlv 420  
 QY 419 PVDYFLSLGVLWLGKFGITSAFSLMXYFTAEPLTVLRNMAVGVTSTASRGSIIAPYF 478  
 DB 421 pbdlylatvlnvvgkfgtaafsmvvytaelyptvtrnmvgvgsstasrsglsilspyf 480  
 QY 479 VYLGAYNRLPLVYVNGSLTVLIGITFTFFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535  
 DB 481 vylgaydrflpylmgstiltaltiflpsfgtpltidqmlrvkgmkkrktpshtr 540  
 QY 536 DSMETEENPKVL-ITAF 551  
 DB 541 mldgqerptlikstaf 557

RESULT 5  
 ID AAY83929 standard; Protein; 557 AA.  
 XX  
 AC AAY83929;  
 XX  
 DT 05-JUL-2000 (first entry)  
 XX  
 DE Human carnitine transporter protein OCTN2.  
 XX  
 KW Organic cation transportation; human; carnitine transporter protein;  
 KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;  
 KW Juvenile visceral steatosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200014210-A1.

XX 16-MAR-2000.  
 PD  
 XX 07-SEP-1999; 99WO-JP04853.  
 PF  
 XX 07-SEP-1998; 98JP-0252683.  
 PR  
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 PA  
 XX Nezu J, Oku A;  
 PI  
 XX WP1; 2000-256966/22.  
 DR  
 DR N-PSDB; AAA09889.  
 XX  
 PT Systemic carnitine deficiency gene OCTN2 encoding part of organic  
 PT cation transporter, useful as diagnostic tool  
 XX  
 PS Example 1; Page 36-40; 106pp; Japanese.  
 XX  
 CC This sequence represents the human carnitine transporter protein OCTN2.  
 CC The coding sequence can be used as a target for diagnosis of systemic  
 CC carnitine deficiency by detecting the presence of mutations in the  
 CC sequence, especially seen in the disease juvenile visceral steatosis  
 CC (jvs). The wild type OCTN2 gene can be used in the gene therapy of the  
 CC disease state.  
 XX  
 SQ Sequence 557 AA;  
 Query Match 78.0%; Score 2218; DB 21; Length 557;  
 Best Local Similarity 75.9%; Pred. No. 1.5e-217;  
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

QY 1 MRDYDEVIAFLGEMGPFORLFFLLSASIPNGFNGMSVFLAGTPEHRCRVPDANLSS 60  
 DB 1 mrdydeviaflgawpffgrllffllsasiipngftgssvflattpehrcrvpdaanlss 60  
 QY 61 AWNNNSVPLRLRGREVPSCSRVRLATIANFSALEPGRDVDLQLEQESCLDGEWFS 120  
 DB 61 awnhtvplrlrdgrevphscrryrlatiansaalgldpgrdvdldglegescldgweifs 120  
 QY 121 QDYLSTVTVTEWNLNCEWNVPLTTSILFFVGLGSEFVSQSLSDRFGKKNVLEATMAVQ 180  
 DB 121 qdylstivtewnlvnceddwkapitlsilffvgvllgsfsgqslsdrgfrknvlftmgnmq 180  
 QY 181 TGFSLQIFSIWENFTVLFVIVGMOISNVYVAFILGTEILGKSVRIIFSTLGCTFFA 240  
 DB 181 tgsfllqifsknfemfvvlvlgmgqisnyvaafvlgteilgksvriifstlgvcifya 240  
 QY 241 VGMULLPLFAYEIRDRWMLLALAVPGVLCVPLWVWTFESPRWLSIORFREAEITQKA 300  
 DB 241 fgmvlplfayfirdwrmllalavpgvllcvplwvwtvfpesprwlsiorsrfreaeditqka 300  
 QY 301 AKNNNTAVPAVIFD--SVEELNPLKQKAFILDLFRTNIAIMTINSLMLMTSVGYFA 358  
 DB 301 akangivvpstifdpseqlqdlsskkqgshnldlrltwnlrmvtimsimlwtisvgyfg 360  
 QY 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAAVFLWGGVGLFIQIV 418  
 DB 361 lsldpnlhgdifvncflsalamvepayvawlllqylprysmatalfggavllfmqlv 420  
 QY 419 PVDYFLSLGVLWLGKFGITSAFSLMXYFTAEPLTVLRNMAVGVTSTASRGSIIAPYF 478  
 DB 421 pbdlylatvlnvvgkfgtaafsmvvytaelyptvtrnmvgvgsstasrsglsilspyf 480  
 QY 479 VYLGAYNRLPLVYVNGSLTVLIGITFTFFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535  
 DB 481 vylgaydrflpylmgstiltaltiflpsfgtpltidqmlrvkgmkkrktpshtr 540  
 QY 536 DSMETEENPKVL-ITAF 551  
 DB 541 mldgqerptlikstaf 557

RESULT	6	
AA01652		
ID	AA01652 standard; Protein; 557 AA.	
AC	AA01652;	
XX		
XX		
DT	23-JUN-1999 (first entry)	
DE	A protein with cation transporting activity.	
XX		
XX	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;	
KW	heart disease; cancer; anti-tumour drug; anticancer drug.	
KW		
XX		
OS	Mus musculus.	
XX		
PN	W09913072-A1.	
XX		
PD	18-MAR-1999.	
XX		
PF	07-SEP-1998; 98WO-JP04009.	
XX		
XX	20-MAY-1998; 98JP-0156660.	
PR	08-SEP-1997; 97JP-0260972.	
XX		
XX	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.	
PA		
XX	Nezu J, Oku A;	
PI		
XX		
XX	WPI; 1999-215062/18.	
DR	N-PSDB; AAX26902.	
XX		
XX	Genes homologous with organic cation transporters OCT1 and OCT2,	
PT	useful in design of new drugs for treatment of diseases due to	
PT	abnormality of the transporter functions	
XX		
XX	Claim 1; Page 75-79; 97pp; Japanese.	
PS		
XX		
CC	The present sequence represents a protein with cation transporting	
CC	activity. The genes are significantly homologous with organic cation	
CC	transporters OCT1 and OCT2. The genes may be used in drug development,	
CC	particularly in the treatment of diseases due to abnormality of the	
CC	organic cation transporter functions e.g. fatty liver, heart diseases	
CC	and cancers, by controlling such as by inhibition or activation.	
CC	Administration of anti-tumour and anticancer drugs in combination with	
CC	a transporter protein inhibiting agent allows the agents to penetrate	
CC	into the diseased cells to enhance the drug action.	
XX		
SQ	Sequence 557 AA;	
Query Match 76.1%; Score 2164; DB 20; Length 557;		
Best Local Similarity 72.9%; Pred. No. 4.9e-212;		
Matches 406; Conservative 70; Mismatches 75; Indels 6; Gaps 3;		
QY	1 MRDYDEVIAFLGEGPQFORLIFFLLSASITPENGFMGMSVFLAGTPBRCRVPDAANLSS 60	
DB	1 mrdydeviaflgwgppqrlifllsasiipngfngmsivflagtpbrcrlvptvnls 60	
QY	61 AWRNNSVPLRLDGRVPHSCSRRLATIANFSAIGLEPGRDVDLGLESCLDGWEFS 120	
DB	61 awrnhsipletkddgrvpqkerryrlatianselglepgrdvdlqegescldgweyd 120	
QY	121 ODVYLSTVVTWNLCENWKPVLTTSLFFGVLLGSFVSGQLSDRFRGNVLFATMAVQ 180	
DB	121 kdvflstvtwvdlvckdwdkapitlslffvgvlmgfsiqglsdrgrknvfltmq 180	
QY	181 TGFSFLOIFSISWEMFTVLFVVGMGQTSYVAVFILGTTEILGKSVRIIPSTLGVCTFFA 240	
DB	181 tgfsflgfvsvnfemftvlflvlgmgqisnyvaafvlgteillsksirliatlgvcifya 240	
QY	241 VGYMLLPFAFIRDRWMLLALTPGVLCVPLWVFPESPRLWISQRRFREAEDIIQKA 300	
DB	241 vgmllpfafrdrwmlallaltvgvlgcvplwvfpesprwlisqrrfreakediioka 300	

```

Query Match          76.1%; Score 2164; DB 21; Length 557;
Best Local Similarity 72.9%; Pred. No. 4.9e-212;
Matches 406; Conservative 70; Mismatches 75; Indels 6; Gaps 3;

Qy 1 MRDYDEVIAFLGEGWGPORLIFFLLSASIIIPNGFNGMSVFLAGTPEHRCRVPDAANLSS 60
Db 1 mrdydeviaflgwgppqrlifllsasiipngfngmsvflagtpchrcrlvphvtnlss 60
Qy 61 AWRNNSVPLRLRDGVRPHSCSRRLATIANFSALEGRDVLQLEQESCLDGEWFS 120
Db 61 awrnhsipletkdgrvqpkrryrlatianselglepdrvdleqlescldegweyd 120
Qy 121 QDVYLSVTWVWNLVCEENKVKPLTSLFFVGLGSFVSQGLSDRFGKKNVLFATMAVQ 180
Db 121 kvdfstivtewdlvckdwdkapltslffvglmgsfvsqglssdrfgrknvlfcmgmq 180
Qy 181 TGFSFLQIFSIWEMFTVLFVIVGMQISNVVAFILGTILGKSVRIIFSTLGVCTFFA 240
Db 181 tgfsflqvsvnfemftvlfvlgmgqisnyvaafvlgteillsksiriffatlgvcifva 240
Qy 241 VGYMLLPFAVFIKDWMLLALVPGVLCVPLWFWFIPESPRWLISORRPREAEDIIOKA 300
Db 241 fgfmvlpfayfirdwrmllaltvpgvlgcalwffipesprwlisgriakeavirka 300
Qy 301 AKMNTAVPAVIFD--SVEELNPLKQOKAFILDLFRTNIAIMTMSLLMLTSSVGYFA 358
Db 301 akingivapstifdpseqlqinstkplhhiydlirtnirvritmsillwlcisvgyfg 360
Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYITAAVLFWGGVLLFTQLV 418
Db 361 lsldcpnlhgdlyvncflaavepayvawlllqyiprrysisaalflgssvllfmqlv 420
Qy 419 PVDYVFLSIGLVMLGKFGITSAFSLMYVFAELVPTLVNRNMAVGVSTASRVGSIIAPYF 478
Db 421 pselftyistalvmgkfgitsaysmvvytaelyptvvrnmvgvsstaarlsgilspyf 480
Qy 479 VYLGAYNRMPLPYVNGSLTVLIGITFLFFPESLGMTLPETLEOMOKVKWPRSKG---KTR 535
Db 481 vylgaydrflpylmgsltlaltlffpesfgvpdpdtldqmlrvkgikqwgqsgqr 540
Qy 536 DSMETEENPKVL-ITAF 551
Db 541 mqkdggeesptvltkstaf 557

RESULT 8
AAY83930
ID AAY83930 standard; Protein: 557 AA.
XX
AC AAY83930;
XX
DT 05-JUL-2000 (first entry)
XX
DE Mouse carnitine transporter protein OCTN2.
XX
KW Organic cation transportation; mouse; carnitine transporter protein;
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
KW juvenile visceral steatosis.
OS Mus musculus.
XX
PN WO200014210-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-UP04853.
XX
PR 07-SEP-1998; 98JP-0252683.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX

```

```

PI Nezu J, Oku A;
XX
DR WPI: 2000-256966/22.
DR N-PSDB: AAA09890.
XX
PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
PT cation transporter, useful as diagnostic tool.
XX
PS Example 1; Page 47-51; 106pp; Japanese.
XX
CC This sequence represents the mouse carnitine transporter protein OCTN2.
CC The coding sequence of the corresponding human protein can be used as a
CC target for diagnosis of systemic carnitine deficiency by detecting the
CC presence of mutations in the sequence, especially seen in the disease
CC juvenile visceral steatosis (jvs). The wild type OCTN2 gene can be used
CC in the gene therapy of the disease state.
XX
SQ Sequence 557 AA;

Query Match          76.1%; Score 2164; DB 21; Length 557;
Best Local Similarity 72.9%; Pred. No. 4.9e-212;
Matches 406; Conservative 70; Mismatches 75; Indels 6; Gaps 3;

Qy 1 MRDYDEVIAFLGEGWGPORLIFFLLSASIIIPNGFNGMSVFLAGTPEHRCRVPDAANLSS 60
Db 1 mrdydeviaflgwgppqrlifllsasiipngfngmsvflagtpchrcrlvphvtnlss 60
Qy 61 AWRNNSVPLRLRDGVRPHSCSRRLATIANFSALEGRDVLQLEQESCLDGEWFS 120
Db 61 awrnhsipletkdgrvqpkrryrlatianselglepdrvdleqlescldegweyd 120
Qy 121 QDVYLSVTWVWNLVCEENKVKPLTSLFFVGLGSFVSQGLSDRFGKKNVLFATMAVQ 180
Db 121 kvdfstivtewdlvckdwdkapltslffvglmgsfvsqglssdrfgrknvlfcmgmq 180
Qy 181 TGFSFLQIFSIWEMFTVLFVIVGMQISNVVAFILGTILGKSVRIIFSTLGVCTFFA 240
Db 181 tgfsflqvsvnfemftvlfvlgmgqisnyvaafvlgteillsksiriffatlgvcifva 240
Qy 241 VGYMLLPFAVFIKDWMLLALVPGVLCVPLWFWFIPESPRWLISORRPREAEDIIOKA 300
Db 241 fgfmvlpfayfirdwrmllaltvpgvlgcalwffipesprwlisgriakeavirka 300
Qy 301 AKMNTAVPAVIFD--SVEELNPLKQOKAFILDLFRTNIAIMTMSLLMLTSSVGYFA 358
Db 301 akingivapstifdpseqlqinstkplhhiydlirtnirvritmsillwlcisvgyfg 360
Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYITAAVLFWGGVLLFTQLV 418
Db 361 lsldcpnlhgdlyvncflaavepayvawlllqyiprrysisaalflgssvllfmqlv 420
Qy 419 PVDYVFLSIGLVMLGKFGITSAFSLMYVFAELVPTLVNRNMAVGVSTASRVGSIIAPYF 478
Db 421 pselftyistalvmgkfgitsaysmvvytaelyptvvrnmvgvsstaarlsgilspyf 480
Qy 479 VYLGAYNRMPLPYVNGSLTVLIGITFLFFPESLGMTLPETLEOMOKVKWPRSKG---KTR 535
Db 481 vylgaydrflpylmgsltlaltlffpesfgvpdpdtldqmlrvkgikqwgqsgqr 540
Qy 536 DSMETEENPKVL-ITAF 551
Db 541 mqkdggeesptvltkstaf 557

RESULT 9
AAB20578
ID AAB20578 standard; Protein: 564 AA.
XX
AC AAB20578;
XX
DT 11-DEC-2000 (first entry)
XX

```

DE Mouse OCTN3 protein SEQ ID NO:1.  
 XX Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;  
 KW Identification; regulator; carnitine transport.  
 XX Mus musculus.  
 OS  
 XX WC2000046368-A1.  
 PN 10-AUG-2000.  
 XX  
 XX 04-FEB-2000; 2000WO-JP00619.  
 XX 05-FEB-1999; 99JP-0028406.  
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX Nezu J, Ose A;  
 XX WPI: 2000-586982/55.  
 XX N-PSDB; AAA88053.  
 XX Organic cation transporter gene OCTN3 expressed in testis for  
 PT identification of regulators of carnitine transport for use as drugs  
 XX  
 XX Claim 1; Page 34-39; 58pp; Japanese.  
 XX The present invention describes a mouse organic cation transporter  
 CC protein (OCTN3). Also described are: (1) a method for screening  
 CC compounds for their ability to regulate the transport of an organic  
 CC cation into the cell, by generating a cell expressing OCTN3 at the  
 CC cell membrane, contacting with the compound and organic cation, and  
 CC observing the degree of transport of the organic cation; and (2) a  
 CC method for screening compounds for their ability to be transported into  
 CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell  
 CC membrane, contacting with the compound and observing the degree of  
 CC transport of the compound. OCTN3 can be used for the identification of  
 CC regulators of the transport of organic cations (especially carnitine)  
 CC into cells by OCTN3, for use as drugs. The present sequence represents  
 CC mouse OCTN3.  
 XX  
 SQ Sequence 564 AA:  
 Query Match 71.5%; Score 2034.5; DB 21; Length 564;  
 Best Local Similarity 69.08; Pred. No. 8 3e-199;  
 Matches 380; Conservative 75; Mismatches 85; Indels 11; Gaps 3;  
 QY 1 MRDDEVIAFLGEGPQFORLIFLLSASIIIPNGFNGSVFLAGTPEHRCRVPDAANLSS 60  
 DB 1 mldydevtaflgwgftgrllfillsasilpngftgisavfitaiphehrcrptdvnlss 60  
 QY 61 AWRNNSVPLRLRDGEVPHSCSRVRLATIANFSALGLEPGRDVDLGOLESCLDGWEFS 120  
 DB 61 awrnhsipmetkddgevpqkrryrlatianfselgpegrdvldleqencldgweyd 120  
 QY 121 ODVYLSVTWVTEWNLVDENWVPLTTSFFVGLVSGVSGQLSDRFGRKNVLPATMAVQ 180  
 DB 121 kdfistivtedwlvckdwkaplttsffvvgvllgsfsgqlsdrgrknllftmah 180  
 QY 181 TGFSELOIFISWEMFTLVFVVGQISNYVVAFLIGTEILGKSVRIIPSTLGVCTFFA 240  
 DB 181 tgfisqifvsnfemftllylvgmghisnyvaafvlgtemlsksvrliaftlglvciffa 240  
 QY 241 VCYMLLPFAFIRDRWMLLALVPGVLCVPLWFWFIPESPRWLLSQRRRFEADITQKA 300  
 DB 241 tgfmlvlpfayfirwrrllailtpgvlcalwfwfipesprwllsqrrrfeaditka 300  
 QY 301 AKMNNATPAVIFDSVEELNPL-----KQKAFILDLFRNTAINTMSLLWMLTSV 354  
 DB 301 akngivapstfip-setnklqddskpshhdydvrcpnrilimsiilwltisv 359  
 QY 355 GYFALS LAPN LHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYITAAVLFWGGVLLF 414

Db 360 gyfglsldtpnlnngnyvncflaavepayvlawlllqhvsrrysmagsiflgsvlll 419  
 QY 415 IQLVPDYVYFISGLVLMGLKFGITSAFSMLYVFTAEIYPTLVRNMAVGVTSASRVGSII 474  
 Db 420 vqlvpsdihyistllvmvgkfgitsaysmvvytaelyptvrmvgvsvstasrlgsil 479  
 QY 475 APFYVLGAYNRMLPYIVMGSLSLTVLIGITFTFFPESLGMTLPETLEOMQKVKWFRSGRKT 534  
 Db 480 spfyvlgaydrllpyllmgslltaltitffpessgyslpetidemqkvkkl----kq 535  
 QY 535 RDSMETEENPK 545  
 Db 536 rqskskkgspk 546

RESULT 10  
 AAB43038  
 ID AAB43038 standard; Protein; 560 AA.  
 XX  
 AC AAB43038;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF2802 polypeptide sequence SEQ ID NO:5604.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antipariatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach M;  
 XX WPI: 2000-602362/57.  
 DR N-PSDB; AAC7747.  
 DR  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 XX Claim 11; Page 4791-4792; 5507pp; English.  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antipariatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

QY	54	-----DAANLSAARNNSVPLRLRGREVPHS-CSRYRLATIANFSALGLEPRDV	103
Db	70	hnhnswnledtgallsagqdyvtvgllgellwelscrnkr-----entsslyge	121
QY	104	DLQLQESCLDGEFSSQDYVLTVVTEWNLYCEDNWKVPLTTSFFVGVLGSGFVSGL	163
Db	122	ytgskkfpcvdgylqdygtwtkstavgtnwlvcdrkwlamlqplfmfvllgsvtfgy	181
QY	164	SDRFGKKNVLFATMAVOTGFSFLOIFSISWEMFTVLFTVVGNGOISNYVVAFLITGEIIG	223

```

Db 182 sdrlgrrvltwssmflgfaafadytftmaarfllamvasgylvvgfvyvmefig 241
QY 224 KSVRIIFSLGCTFFAVGYMLLPFLFAYFIRDW----RMLLLALTPVGLVPLMWFIPES 280
Db 242 mksr-twasvhlhsffavgtllvaltgylvrtwlygmilstvtvpfilcc---wvipet 297
QY 281 PWWLSQRFRAEDIIQKAAKNWTA-----VPAVIFDSVEELNPLKQKAFILDLFR 334
Db 298 pfwllsegyeaaqkivdimakwnrasscklsellidlgpvsnsptevakhnllylfy 357
QY 335 TRNIAMTMSLLMLLWLTSGVGFALSDDAPNLHGDAYLNCFLSALIEIPAYITAWMLLRT 394
Db 358 nwsikrtltvliwltfsglsgfyslnsvnlgnneylnlflggvveipaytfvciamdk 417
QY 395 LPRYITIAAVLFWGGVLLFIOLVPDYFLLSIGLWGLK-FGITSAPSMLYVFTAEIYP 453
Db 418 vgrtviaylsfcsalacgvnmvlpqkhyilgvvtamvgkilpigaafglylytaelyp 477
QY 454 TLVRNMAVGVTASRVGSIAPFYVYLGAYNKMPLPYVMGSLTVLIGITLFFFPESLGM 513
Db 478 tivrtiavsgsmvrcilaipfsvdlsiwifipqlfvgtmallsgvltklpetlgk 537
QY 514 TLPTEQMQKWKFRSGKTKTRDSMETENPKVLT 549
Db 538 rlatwteeaakl-----eseneskskillt 563

RESULT 12
AAB49401
ID AAB49401 standard; Protein; 540 AA.
AC AAB49401;
XX
DT 02-MAR-2001 (first entry)
DE Murine organic anion transporter 6.
KW Murine; organic anion transporter 6; mOATP6; cancer; inflammation;
KW cardiovascular disease; central nervous system disorder; kidney disease;
KW liver disease; autoimmune disease.
XX Mus sp.
XX W0200070048-A1.
XX 23-NOV-2000.
XX 15-MAY-2000; 2000WO-US13316.
XX 14-MAY-1999; 99US-0134137.
XX 12-MAY-2000; 2000US-0570293.
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX Feild J, Yue L, Ellens H;
XX
XX WPI; 2001-016235/02.
XX N-PSDB; AAC83979.
XX
XX Murine organic anion transporter 6 polypeptide, useful for identifying
XX agonists/antagonists that are useful in treatment of cancer, kidney
XX disease, autoimmune disease, inflammation and cardiovascular disease -
XX
XX Claim 2; Page 28-29; 32pp; English.
XX
XX The present sequence is murine organic anion transporter 6 (mOATP6).
XX mOATP6 protein is useful for screening compounds which inhibit or
XX stimulate the function of mOATP6 and also compounds that neither agonise
XX nor antagonise OATP6. The identified agonists and antagonists are useful
XX for prevention and treatment of human diseases, including cancer,
XX inflammation, cardiovascular disease, central nervous system disorders,

```

CC kidney diseases, liver disease and autoimmune diseases.

XX Sequence 540 AA;

Query Match 24.5%; Score 698; DB 22; Length 540;  
 Best Local Similarity 32.5%; Pred. No. 1.9e-62;  
 Matches 183; Conservative 97; Mismatches 231; Indels 52; Gaps 19;

```

QY 4 YDEVIAFLGEGWGPQ--RLIFFLLSALIIINGNGSVVFLAGTPEHRCRVPDA-ANLS- 59
Db 3 feelihkvvgfgrqlnvlalprflipmhf--lipfmaavpahhcaipanalsh 60
QY 60 -SARNNSVPLRLRDGREVPVHSCSRYLATIANFSALGLEPGRDVLGLQEQE-----SCL 114
Db 61 qdlwlkthlp-retdgsf--ssclrfaypqlpnvtlgtv---ynsgepegepltpcs 114
QY 115 DGWEFSODVYLSTVTWENLVCEENWKVPLTSLTFFVGVLLGSFVSQLSDRGKKNVLF 174
Db 115 qgweydrsefsstiatcdwlcqgrlnkvtstcffiigllgavvyeylsdrfgrirll 174
QY 175 ATMAVQTGFSFLQIFISWEMFTVLFVIVGMQISNVVAFILGTEILGKSVRIIFSTLG 234
Db 175 vaystlalgimsaasvnylmfvtrmltsalagftliivlplelewldehrtvagvis 234
QY 235 VCTFFAVGYMLLPFLFAYFIRDRMLLLALTPVGLVCPV---LWMFIPESPRMLISORRF 290
Db 235 t-tftgvgvllltlvgylirswrwillaatlp---cypgiisiww--vpesarwlltqgrv 289
QY 291 REAEDIITQAKAMNNTAVPAVIFDSV--EELNPL-----KQKAFILDLFTRNIAIMI 343
Db 290 eekkyisicaklngripse---dlsqsalnkvtmervsqrpsylidfrtsqlrhvsl 346
QY 344 MSLLLWMLTSVGYFALSDDAPNLHGDAYLNCFLSALIEIPAYITAWMLLRLTPRR---- 398
Db 347 commmwfgnfsyvgltldasgltyqtllfgavevpskitvfflvrlvgrlreag 406
QY 399 YIIAAVLFWGGVLLFIOLVPDYFLLSIGLWGLKFGITSAPSMLYVFTAEIYPVLVRN 458
Db 407 mllatalctfgisl-----lvssdtkswitalvlgkafseaaftaylftselypvlrq 461
QY 459 MAVGVTSASRVGSIAPFYVYLGAYNKMPLPYVMGSLTVLIGITLFFFPESIGMTLPET 518
Db 462 tgmgtfalgigrigaslapvllldgwwlllpklayggisflaactvlllpetkkaipet 521
QY 519 LEQMOKVKWFRSGKTKTRDSMETE 541
Db 522 lqdv-----rkgrki-drsqte 538

RESULT 13
AAB64538
ID AAB64538 standard; Protein; 554 AA.
XX
XX AAB64538;
XX
XX 21-OCT-1998 (first entry)
XX Human liver cell clone HP01293 protein.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
XX differentiation; immune system; stimulator; suppressor; regulator;
XX hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
XX haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
XX Homo sapiens.
XX
XX W09821328-A2.
XX
XX 22-MAY-1998.
XX
XX 07-NOV-1997; 97WO-JP04056.
XX

```

```

PR 13-NOV-1996;      96DP-0301429.
XX
PA (PROT-) PROTEGENE INC.
PA (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX
WPI: 1998-297932/26;
DR N-PSDB: AAV49558, AAV49559.
XX
Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX
PS Claim 1; Page 96-98; 205pp; English.
XX
AAW64534-W64558 represent human proteins containing a transmembrane
CC domain. These proteins can be used for, e.g. research and nutrition, and
CC may have cytokine and cell proliferation/differentiation, immune
CC stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activity.
XX
SQ Sequence 554 AA;

Query Match          24.1%; Score 685.5; DB 19; Length 554;
Best Local Similarity 33.0%; Pred. No. 3.8e-61;
Matches 190; Conservative 87; Mismatches 229; Indels 69; Gaps 14;

QY 1 MRDYDEVTAFLGECWGPFOR---LIPFLLASATIPNGMSVVFLAGTPHEHRCRVPDAAN 57
Db | : : : : | | | : | | | : | | | : | | | : | | | : | | | : | | | : |
QY 1 mptvdvdlleqvgesgwfkdqafllcllisaafap---fcvgivflgfptdhhcqspgvae 57
Db | : : : : | | | : | | | : | | | : | | | : | | | : | | | : | | | : |
QY 58 LSS---AWR-----NNSVPLRLRDGREVPUSCSRYR-----LATIA-NFSALG 96
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 lsqcgwspaeelnlytvpqlpageaflgqrryevdwnqsalcsvdplasiatnrshlp 117
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 97 LEPRGDVDLGLEEQESCLDGHFEFDVYLSTVTVENLVCEBDNKVKPLTTSLFFVGVLGG 156
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 lgp-----cqdgvwy--dtpgssivtfenlvicadskwldlfqsclnagfffg 162
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 157 SFVSQSLSDFGRKNLFATMAVQTGFSEFLQIFSIISWEMFTVLFVIVGMGOISNWVVAFI 216
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 163 slvgvyfadrfrgkiclgtvlvnavgvmlnafspymsmllflrlggivlskgnmagyt 222
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 217 LGTEILG----KSVRIPTSGTCFTFAVGWMLLPFLPAYFIRDWRMLLLALTLPGLVCVP 272
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 223 litefvsgsrtrvalmyqma-----ftcvgaigtlayalphwrwlqlavslptfil 277
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 273 LWRTIPSPRWLIQSRRFREAEITIQAKANNNTAVPAVI-----PDSVEELNPLQOK 326
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 278 yywcvsprwllsqkrnteaikimdhiaqngklppadikmlesleedvtekisp----- 332
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 327 AFILDLFETRNIATMTIMSLLLMWMLTSVGVPALSDDAPNLHGDAYLNCFLSALIEI PAYI 386
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 333 -sfadlfrtplrktfilmyltwfdsvlygglllhmgatsgnlylglfsalsylveigaf 391
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 387 TAWLLRTLPRXYIIAALFWGGGVLLFIQLVPDYDFLSTGLVMKGFGITSFMSLYV 446
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 392 ialitidrvgrlypmavsnllagaacvlmfisdpdlhlmlmcvgrmgtitaigmicl 451
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 447 FTAELYPTFLVRNMNAVGTSTASTRVGSIIAPFYFV-LGAYNRMLPYIVMGSTVLIGIFTL 505
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 452 vnaelyptfrnlgvmvcsslcdggiltptfivlrrevwgpalpli favllglaegvtl 511
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 506 FFPESLGWLTPELQEWQKVWFERSGKKTRDSMET 540
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 512 llpeckgvaipcmkdaenl-----grkakpkt 541
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
AAV51249
ID AAY51249 standard; Protein; 535 AA.
```



Db	408	ll--gaaltfgtllivslsletkswitalvvgkafseaafttaylftselvptvirqtglg	465
Qy	463	VfSTASRVGSIITAPYFVYLGAYNRLPYIVMGSLNVLICIGFTLFPPEISLGMTLPETLEOM	522
Db	466	ltalmgrlgaslaplaalldgwlilpkvayggiaivaactallilpetckaqilpetiqdv	525
Qy	523	QK 524	
Db	526	er 527	
RESULT 15			
ID	AAR77676		
AC	AAR77676		
XX	10-SEP-1996	(first entry)	
XX	Rat OCT-1 protein.		
DE	Rat; OCR-1; transporter protein; cationic; xenobiotic; pharmaceutical		
KW	blood; liver; kidney; epithelial cell; intestine; tetraethylammonium;		
KW	proximal renal tubule cell; intestine; enterocyte; transgenic; renal;		
KW	biliary; excretion; resorption; modulator; uptake.		
OS	Rattus rattus.		
XX			
Key	Location/Qualifiers		
Domain	20..46		
FT	/note= "transmembrane domain"		
FT	71		
FT	/note= "putative N-glycosylation site"		
FT	97		
FT	/note= "putative N-glycosylation site"		
FT	113		
FT	/note= "putative N-glycosylation site"		
FT	154..171		
FT	/note= "transmembrane domain"		
FT	178..197		
FT	/note= "transmembrane domain"		
FT	243..260		
FT	/note= "transmembrane domain"		
FT	267..283		
FT	/note= "transmembrane domain"		
FT	350..366		
FT	/note= "transmembrane domain"		
FT	380..398		
FT	/note= "transmembrane domain"		
FT	406..425		
FT	/note= "transmembrane domain"		
FT	432		
FT	/note= "putative N-glycosylation' site"		
FT	435..452		
FT	/note= "transmembrane domain"		
FT	469..485		
FT	/note= "transmembrane domain"		
FT	494..514		
FT	/note= "transmembrane domain"		
XX			
PN	DE4424577-A1.		
XX			
PD	18-JAN-1996.		
XX			
PF	13-JUL-1994;	94DE-4424577.	
XX			
PR	13-JUL-1994;	94DE-4424577.	
XX			
PA	(FARH ) HOECHST AG.		
XX	Gorboulev V, Gruendeman D, Koepsell H;		
XX			

Fri Jan 25 10:19:00 2002

us-09-521-195-1.rag

Page 12

Job time: 53 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 10:08:24 ; Search time 23.33 seconds  
(without alignments)  
531.475 Million cell updates/sec

Title: US-09-521-195-1  
Perfect score: 2845  
Sequence: 1 MRDYDEVIAFLGEGWPFQRL.....KKTRDSMETENPKVLITAF 551

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	702	24.7	555	3	US-08-501-572-3
2	702	24.7	555	3	US-09-040-444-3
3	687.5	24.2	556	3	US-08-501-572-1
4	687.5	24.2	556	3	US-09-040-444-1
5	680	23.9	553	3	US-08-501-572-2
6	680	23.9	553	3	US-09-040-444-2
7	630	22.1	537	2	US-08-647-397-2
8	267	9.4	520	4	US-08-964-127-2
9	234.5	8.2	494	2	US-09-031-392-5
10	234.5	8.2	494	4	US-09-299-549-5
11	222	7.8	492	2	US-08-355-844-3
12	222	7.8	492	5	PCT-US95-16126-3
13	216	7.6	493	2	US-09-031-392-10
14	216	7.6	493	4	US-09-299-549-10
15	213.5	7.5	524	2	US-08-928-692-12
16	203	7.1	286	4	US-08-964-127-4
17	196	6.9	500	2	US-09-031-392-7
18	196	6.9	500	4	US-09-299-549-7
19	181	6.4	509	2	US-09-031-392-6
20	181	6.4	509	4	US-09-299-549-6
21	174.5	6.1	584	2	US-08-928-692-13
22	172	6.0	383	2	US-09-031-392-3
23	172	6.0	383	4	US-09-299-549-3
24	169.5	6.0	488	2	US-08-928-692-11
25	168.5	5.9	488	2	US-08-928-692-10
26	162	5.7	563	2	US-09-031-392-2
27	162	5.7	563	4	US-09-299-549-2

28	158.5	5.6	109	2	US-08-647-397-4	Sequence 4, Appli
29	156.5	5.5	534	2	US-09-031-392-4	Sequence 4, Appli
30	156.5	5.5	534	4	US-09-299-549-4	Sequence 4, Appli
31	149.5	5.3	322	4	US-08-964-127-6	Sequence 6, Appli
32	110	3.9	473	1	US-08-597-236-13	Sequence 13, Appli
33	110	3.9	473	1	US-08-746-682A-13	Sequence 13, Appli
34	110	3.9	548	3	US-08-903-139B-9	Sequence 9, Appli
35	109.5	3.8	834	2	US-08-677-734A-9	Sequence 9, Appli
36	109.5	3.8	834	2	US-08-677-734A-10	Sequence 10, Appli
37	108.5	3.8	1299	4	US-08-460-900C-62	Sequence 62, Appli
38	108.5	3.8	1299	4	US-08-674-509B-48	Sequence 48, Appli
39	108.5	3.8	1299	4	US-08-954-698-48	Sequence 48, Appli
40	107	3.8	1285	2	US-08-540-406-6	Sequence 6, Appli
41	107	3.8	1285	3	US-08-656-055-6	Sequence 6, Appli
42	107	3.8	1285	4	US-08-954-668-6	Sequence 6, Appli
43	107	3.8	1285	5	PCT-US95-13233-6	Sequence 6, Appli
44	103.5	3.6	1286	4	US-09-268-140-3	Sequence 3, Appli
45	103	3.6	548	3	US-08-903-139B-28	Sequence 28, Appli

ALIGNMENTS

RESULT 1  
US-08-501-572-3  
; Sequence 3, Application US/08501572  
; Patent No. 6063623  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; APPLICANT: Gorboulev, Valentin  
; TITLE OF INVENTION: Transport protein Which Effects The  
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,  
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/501,572  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toohey, Kimberlin M  
; REGISTRATION NUMBER: 35,391  
; REFERENCE/DOCKET NUMBER: 02481.1453-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 555 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-501-572-3

Query Match 24.7%; Score 702; DB 3; Length 555;  
Best Local Similarity 33.7%; Pred. No. 5.3e-65;  
Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;  
QY 5 DEVIAFLGEGWPFQRLIFF---LLSASIIPIINGFNGMSVYFLAGTPEHRCRVPDANLS-- 59

Db	6	DDVLRGEGEFHFQKQWFFLLALSATP---	IYGVIFGTFDDHCRSPGVAELSRLR	62
QY	60	SAWR-----NNSVRLUROGREVPHSCSYRLA-TIANFSALGLEPGRDVDLQJQJESCL	113	
Db	63	COWSPAELNVTYVGPAGAEAPROCRYEYDWNQSTFDCVPLASLDIUNRSRLPLGCP	122	
QY	114	LDGWFEFSQDVYLSVWTFEMNLVCEDNWKVPLATSLTFYGVLLIGSFVSGQLSDRFGKKNVL	173	
Db	123	RDGWVY--ETPGSSIVTEFNLYCANSWMOLDFOSSVNWVGFIGSMISGYIADREGRKLC	180	
QY	174	FATMAVQCPSEFLQIESISNEMETVLEVIYVGMQISNVAVAFILCTELGKSVRIIFESTL	233	
Db	181	LTTVLINAAAGVLAISPTTWMILPFLTOGLVSKAGMLIGYILITFVGGRY--RIV	237	
QY	234	GVC--TFPAVGYMLPLFAYEIRDRWMLLALTVPGLVCLFWMFIPESRWLIISORRER	291	
Db	238	GIFYQVAKTVGLLVLAGVAYALPHWRMLOFTVALPNFFFLLYVVICIPESRWLIISOKNA	297	
QY	292	EAEIDIQAAKMNNTAVPAVI-----FDSVEELNPLQOKAFILDLFRNTAINTMTS	345	
Db	298	EAIRLIKHTAKNGKSLPASLQRLREETGKKLP-----SFLDLVYRTPQIKKHTMIL	351	
QY	346	LLIWMMLTSVGYEALSLQAPNKHGD-AVLCNCFLSALIEIPAYITAWLLRLTLPRTYIAAV	404	
Db	352	MYNWFSTSVLYGGIIMHM-GLAGDNTYLDFFYSALVEPFAAFMILIIDIRIGRYPAAS	410	
QY	405	LFWGGGVLLFTQLFVVDYFYSIGLYMGLKRGFTISAFSNLXVFTAELYPTILVRNMVGT	464	
Db	411	NMVAGAACLASVFIIPGDQLWKIISCLGRMGITMAYEIVCLVNAELYPTIFIRNLGVHC	470	
QY	465	STASRVGSIIPAYFVILGAYNRM-----LPYIVMGSLTVLIGLITFFPESLGMLTIPET	518	
Db	471	SSMCDIGGITTFPLYI-----RUTNIWEPLMVFVGLVAGGLLALPETGKALPET	525	
QY	519	LSQMQRKVKWFRSGKK	533	
Db	526	IEEAENMQRPKNKE	540	

## RESULT 2

```

1  US-09-040-444-3
2  :
3  : Sequence 3, Application US/09040444
4  : Patent No. 6063766
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: Koepsell, Hermann
9  :
10 : APPLICANT: Grundeman, Dirk
11 :
12 : APPLICANT: Gorboulev, Valentin
13 :
14 : TITLE OF INVENTION: Transport protein Which Effects The
15 :                      Transport of Cationic Xenobiotics and/or Pharmaceuticals,
16 :
17 : TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
18 :
19 : NUMBER OF SEQUENCES: 6
20 :
21 : CORRESPONDENCE ADDRESS:
22 :
23 : ADDRESS: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
24 :
25 : STREET: 1300 I Street, N.W., Suite 700
26 :
27 : CITY: Washington
28 :
29 : STATE: D.C.
30 :
31 : COUNTRY: USA
32 :
33 : ZIP: 20005-3315
34 :
35 : COMPUTER READABLE FORM:
36 :
37 : MEDIUM TYPE: Floppy disk
38 :
39 : COMPUTER: IBM PC compatible
40 :
41 : OPERATING SYSTEM: PC-DOS/MS-DOS
42 :
43 : SOFTWARE: PatentIn Release #1.0, Version #1.30
44 :
45 : CURRENT APPLICATION DATA:
46 :
47 : APPLICATION NUMBER: US/09/040,444
48 :
49 : FILING DATE: March 18, 1998
50 :
51 : CLASSIFICATION:
52 :
53 : ATTORNEY/AGENT INFORMATION:
54 :
55 : NAME: O'Connor, Steven P
56 :
57 : REGISTRATION NUMBER: 41,225
58 :
59 : REFERENCE/DOCKET NUMBER: 2481.1453-01
60 :
61 : TELECOMMUNICATION INFORMATION:

```



```
Query Match      23.9%; Score 680; DB 3; Length 533;
Best Local Similarity 33.4%; Pred. No. 1.1e-62;
Matches 192; Conservative 85; Mismatches 228; Indels 70; Gaps 15;

1 MRDVEVIAFLGEGWGFQR--LIFELLASAIPINGFSGMSVVLAGTPEHRCRVPDAAN 57
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
1 MPVTDDILEQVGESGFHQKFAFIILCLLSAAP---ICVGIVLFGLTFPDHHCSPGVAE 57
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::

58 LSS--AWR-----NNISVPLRLRDGREVPHSCSYRR-----LATIA-NFSALG 96
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
```

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Connor, Steven P  
REGISTRATION NUMBER: 41,225  
REFERENCE/DOCKET NUMBER: 2481.1453-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)408-4000  
TELEFAX: (202)408-4400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-040-444-2

Query Match 23.9%; Score 680; DB 3; Length 553;  
Best Local Similarity 33.4%; Pred. No. 1.le-62;  
Matches 192; Conservative 85; Mismatches 228; Indels 70; Gaps 15;

QY 1 MRDYDEVIAFCEWCFOR---LIFFLLSASIIPIGNGMSVVFVLAGTPEHRCRVPPDAAN 57  
DB 1 MPTVDIILEQGESGWFQKFLIICLSAFAF---ICVGIVFLGTPDHHCOSPGVAE 57  
QY 58 LSS---AWR-----NNSVPLRLRDGREVPHSCSYR-----LATIA-NFSALG 96  
DB 58 LSQCGWSPAELNVTVPCLGPAGEAFLGQCRRYEVDMNQSALSCVDPPLASLATNRSHLP 117  
QY 97 LEPGRDVLGQLEQESCLDGEHFSQDVYLSIVVTWNVLCEDNKKVPLTTSFFVGVLLG 156  
DB 118 LGP-----CQDQWY---DTFGSSIVTEFNLVCADSNKLDLDFQSCLNAGFFFG 162  
QY 157 SFVSGQLSDRFGRKNVLPATMAVQTFGLQIFSISWEMFTVLFVIVGMQISNYVAFI 216  
DB 163 SLGVGYFADRFGRKCLLGLTVLNAVSGVLMAFSPNYSMLLFRLLQGLVSKGNMAGYT 222  
QY 217 LGTEILG-----KSVRIESTLGVCTFFAVGYMLLPFLPAYFIRDMRMLLALTVPGLVCP 272  
DB 223 LITEFVSGSRRTVAIMYQMA-----FTVGLVALTGLAYALPHWRMLQAVLSLTFLELL 277  
QY 273 LWFIPESRWLLSQRPREAEDIQKAAMNNTAVPAVI-----FDSVEELNPKQOK 326  
DB 278 YMCVPESRWLLSQRNTEAKIMDHIAQKNGKLPADLKMLSLEEDVTEKLS- 332  
QY 327 AFTLDLFRTRNIAIMTMSLLMLTSGVGFALSALDAPNLHGDAYLNCFLSALIEPAYI 386  
DB 333 -SPADLFRPLRKRTFILMYLWFTDSVLYQGLIHMGTSGNLYDLFLYSALVEIPGAF 391  
QY 387 TAWMLLRTLPRIYIAAVLFWGGVLLFTQLVPDYFYFLSIGVLMLKFGITSASFSLVY 446  
DB 392 IALITIDRVGRIPMAVSNLLAGAACLVTISP-DLHWLNIIMCVGRMGITIAQMICL 450  
QY 447 FTAEPLTVLRNMAVGTSTASRVSGSIIAPYFY-LGAYNRMLPIYVMGSLAVLIGIFTL 505  
DB 451 VNALPTTFVRNRVWVCSLSDIGIITFFIVFRREVWQALPLILFAVLGLAAGVTL 510  
QY 506 FPPESLGMLTPLEQMOKVKFRSGKTRDSMET 540  
DB 511 LLPETKDALPETMKDAENL-----GRKAKPKENT 540

RESULT 7  
US-08-647-397-2  
Sequence 2, Application US/08647397  
Patent No. 5972702  
GENERAL INFORMATION:  
APPLICANT: Beiter, David R.  
APPLICANT: Brady, Kevin P.  
TITLE OF INVENTION: OSTEOCLAST TRANSPORTER  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/647,397  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Gates, Edward R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: B0801/7048  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 537 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-647-397-2

Query Match 22.1%; Score 630; DB 2; Length 537;  
Best Local Similarity 30.0%; Pred. No. 1.8e-57;  
Matches 165; Conservative 102; Mismatches 245; Indels 38; Gaps 12;

QY 4 YDEVIAFLGEMGFPQRLIFLFSASIIPNGFNGMSVVFVLAGTPEHRCRVPPDAANLSSAWR 63  
DB 3 FSEILDRVSGMGPQYLVHTLALPILGIANHNLQIIFTATPDHCHCRPPNASL-EPW- 60  
QY 64 NNSVPLRLRDGREVPHSCSYR---RLATIANFALGIEGRDVLGQLEOESCLDGEWFS 120  
DB 61 --VLPLG-PNGK--PEKCLRFVHLPLNASLNDTQATEP-----CLDGMWYN 102  
QY 121 QDVYLSVTWVTWNVLCEDNKKVPLTTSFFVGVLLGSFVSGQLSDRFGRKNVLFATMAVQ 180  
DB 103 ST--RDTIVTEWDLVCGSNKLKEMAQSVFMAGILVGGVFGELSDRFGRKPILTWSYLL 160  
QY 181 TGFSELPQIFSISWEMFTVLFVIVGMQISNYVAFILGTEILGKSVRIIFS-TLGVCTFF 239  
DB 161 AASGSSAAFSLSLTYMIFRFLCGCSISGLSTIILNVEWVPTSTRAISSTIIGY- 218  
QY 240 AVGYMLLPFLPAYFIRDMRMLLALTVPGLVCPVLPWFIPESRWLLSQRPREAEDIQK 299  
DB 219 TIGQFILPCLAYAVPQWRNLQLSVSAFFIFSLLSWVPEISIRVLVSGKFSRAKTLQR 278  
QY 300 AAKMN-----NTAVPAVIFDSVEELNPLKQOKAFILDLPTRNIAIMTMSLLMLT 352  
DB 279 VATFNGKKEGEKLTVEELKFNLDITSAR-VKYGLSDFRVSLRRVTFCLSLAWFAT 337  
QY 353 SVGYFALSADAPNLHGDAYLNCFLSALIEPAYITAWMLLRTLPRIYIAAVLFWGGVYL 412  
DB 338 GFAYSLAMGVVEFGVNIYIILQIFGGVDIPAKFTITLISYLGRRITOGFELLILAGVAI 397  
QY 413 LFIQLVPDYFYFLSIGVLMLKFGITSASFSLYVFTAEPLTVLRNMAVGTSTASRVGS 472  
DB 398 LALIFVSSEMOLLRTALAVFGKGLSGSFCFLYTSLEYLTVLROTGMGTSNIWARVGS 457  
QY 473 ITAPYFVYLGAYNRMPLPIYVMGSLVFLIGITFLFPPESLGMLTPLEQMOKVKFRSGK 532  
DB 458 MIAPLVKITGELQEPINPVNVIEWMTLLGSSAAFFLLETINRPLPETIEDIQ-DWYQOTK 515  
QY 533 KTRDSMETEE 542  
DB 516 KTKQEPKAEK 525

RESULT 8  
US-08-964-127-2  
Sequence 2, Application US/08964127  
Patent No. 6277565  
GENERAL INFORMATION:  
APPLICANT: Grandearl, Andrew David John  
TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE  
MOLECULES

NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/964,127  
FILING DATE: 06-NOV-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crews, Ph.D., L. Lee  
REGISTRATION NUMBER: P-43,567  
REFERENCE/DOCKET NUMBER: 07334/038001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 520 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: Internal  
US-08-964-127-2

Query Match 9.4%; Score 267; DB 4; Length 520;  
Best Local Similarity 23.4%; Pred. No. 2e-19;  
Matches 117; Conservative 88; Mismatches 203; Indels 92; Gaps 21;

QY 85 RLATIANFSAIGLRGRVDLQGEQSCLDGWEFSQDVLST-VVTENWLVCEWKNKVP 143  
DB 50 RVATSDPSCGFAPP-----DFNHCKDWDYGLPVLTTNIGAWDLVCDLQWQVI 101  
QY 144 LFTSLFFVGVLLGFSVGSQSDREGRKNVLFATMAV-----QTGFSFLQIFS 190  
DB 102 LQQLILFGLFASGYLFLGYPADREGRRGIVLLTLGLVPCGVGGAAGSSTGYMALR--- 158  
QY 191 ISWEMFTLVFVGVGQISNVVAFILGTEILGKSVRIIFSTLGVCTFFFAVGVMPLPFA 250  
DB 159 -----FLGLFLLAGVD-----LGVLMRLCQDPTQLRVALAGELVGVGGHFLFLGL-A 207  
QY 251 YPIRDMRLMLLALTVPGVLCVPLMW--FIPESPRWLISORRFEAEIIOKAAMN----- 304  
DB 208 LVSKDWRFLQRMITAPCLFLFYGPGLFLESARWLVKQIEAQSVLRILAEARNRPHG 267  
QY 305 -----NTANPAVIFSVSEINPLKQKAF-ILDLPRTNIAIMTMSLLWMTS----- 353  
DB 268 QMLGEEAQEAL-QDLENTCLPLATSSFSFASCLNRYNI-----WKNLLILGFTNFTAHAI 321  
QY 354 -----VG-----YFALSIDAPNLHGDAYLNC-FLSALIEIPAYITAWLLRLTPRY 399  
DB 322 RHCYQPVGGGSPDFYLCSLA---SSTAALACVFLG--VTVDREGRCILLSWTLG 376  
QY 400 ITAAVLFVGGGVLLFLQLVPVDYVFLSIGLVMLGKFGITSAFMSLVYFPAELYPTLVNRM 459  
DB 377 TASLVLL---GLWLYLNEAAITTF-----SVLGLFSSQAAAILSTLLAAEVIPTTVRGR 427  
QY 460 AVGVSTASRVSITAP-YFVYLCAYNMLPYIVMGSULTVILGIFLTFPESLGMTLPET 518  
DB 428 GLGLNALGALGGLSGPAQRHMG-HGAFLQHVVLACALLCILSINMLPETKRLLPV 486

QY 519 L---EQMKYKWFERSGKKTR 535  
DB 487 LRDGELCRPSLLRQPPPTR 506

RESULT 9  
US-09-031-392-5  
Sequence 5, Application US/09031392  
Patent No. 5942398  
GENERAL INFORMATION:  
APPLICANT: Tartaglia, Louis A.  
APPLICANT: Weng, Xun  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES  
ENCODING GLUTEX AND USES THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/031,392  
FILING DATE: 26-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Meiklejohn, Ph.D., Anita L.  
REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 07334/072001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-031-392-5

Query Match 8.2%; Score 234.5; DB 2; Length 494;  
Best Local Similarity 25.2%; Pred. No. 4.8e-16;  
Matches 115; Conservative 76; Mismatches 156; Indels 109; Gaps 23;

QY 133 NLVCEWKNKVP-----LTTSLFFVGVLLGFSVGSQSDREGRKN--VLFATMAV 179  
DB 43 NYTLERSETPPSSVLLTSLWSLSVAIFSVGGMIGFSVGLFVNRGRNSMLIVNLLAI 102  
QY 180 QTG--FSLQIFSTISWEMFTLVFVIGM--GQISNVVAFILGTEILGKSVRIIFST--- 232  
DB 103 AGGLMGECKI-AESVEMILGLRIIGLFCGLCTGFVPMVI--GEISPTALGAGFTLNQ 159  
QY 233 LGVCTFTFANGVMPLLPAYFI---RDWRMLLALTV-PGVLCVPLMWFPESPRMLISQ 287  
DB 160 LGI---VIGILVAIFGLKVLGTEDLWPLLLGFTLPAIIQCAALPFCSPESPRFLIN 215  
QY 288 RFRFE-----AEDIIQKAAKNNTAVPAVIFDSVEELNPLKQKAFILDLPFR 336  
DB 216 RKEEKAKELQRLWNGEDVAQIQENKQ-----ESMRMSQEKQVTVLELFRAP 264  
QY 337 N-----IAIMTMSLLWMTLSVGYFALSIDAPNLHGDAYLNCFLSALIEIPAYIT--- 387  
DB 265 NYRQPIIISIMQLSQSLGGINAVFYSTGI-----PKDAGVQEPVATIGA 311  
QY 388 -----AWLLRLTPRYITIAAVLFWGG-----GVLLFTQLVPVDYI-----ELSIG- 428



Db 312 GVNNTFTVSVFLVERAGRRTL--HLIGLGMAFCSILMTISLLLLKDNYSWMSFICIGA 369  
QY 429 -LVMLGKGITSAFSLMVLVFAEYPTLVNMAVGVSTAS-----RVGSIAPFYVYL 482  
Db 370 ILVFAVFEIGPG-PIPWFIWAELFGQGPRAAMAVAGCSNWTNLSVLGLLFPSSATEYLG 428  
QY 483 AYNRMPLPIVMSLTVLIGITLFP-PPESLGMTPLE 517  
Db 429 AY-----VFIVTVFLVIFWVTFEKKVPETGRTFEE 460

RESULT 10  
US-09-299-549-5  
; Sequence 5, Application US/09299549  
; Patent No. 6136547  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Beng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/299,549  
; FILING DATE: 26-APR-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/031,392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 494 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-299-549-5

Query Match 8.2%; Score 234.5; DB 4; Length 494;  
Best Local Similarity 25.2%; Pred. No. 4.8e-16;  
Matches 115; Conservative 76; Mismatches 156; Indels 109; Gaps 23;

QY 133 NLVCDNKNVP-----LTSLFFVGVGLIGSVGQLSDRFRGKN--VLFAFMAY 179  
Db 43 NYTLERSETPSSVLLTSLWSLSVAIFSVMGIGSVGLFVNFRGRRNMLIVLLAI 102  
QY 180 QTG--FSPLOFTSWMFTVFLVIVGM--GQISNYVAVILGTEILGKSVRIEST--- 232  
Db 103 AGGCLMGFCIK-ABSVEMLILGRLLIIGLCGLCTGTFPMTYI--GETISPTALRGAFGLNQ 159  
QY 233 LGVCTTFVAVGYMLPLFAEYF-----RWRMLLALTVPGLVCLVPLMWFTIPESPRWLISQ 287  
Db 160 LGI-----VIGILVAQIFGLKVLIGTEDLWPLLGLFTILPAIQCAALPFCPSPRFLIN 215  
QY 288 RRFRE-----AEDIIQAKAMNTAVPAVIFSDSVEELNPLKQOKAFILDLPRT 336

Db 216 RKEERAKELLQRLWGTEDVAQDIQEMKD-----ESMRMSQEQKQVTVLELFRAP 264  
QY 337 N-----IAIMTMSLLMLLTVSGVGFALSLDAPNLHGDAYLNCFLSALIEIPAYIT--- 387  
Db 265 NYRQPIIISINLQSLGAINAVFYSTGI-----FKDAGVQEPVATICA 311  
QY 388 -----AWLLRLTPRRYIIAAVLEWGG-----GVLLFIQLVDPVYI-----FLSIG- 428  
Db 312 GVNNTFTVSVFLVERAGRRTL--HLIGLGMAFCSILMTISLLLLKDNYSWMSFICIGA 369  
QY 429 -LVMLGKGITSAFSLMVLVFAEYPTLVNMAVGVSTAS-----RVGSIAPFYVYL 482  
Db 370 ILVFAVFEIGPG-PIPWFIWAELFGQGPRAAMAVAGCSNWTNLSVLGLLFPSSATEYLG 428  
QY 483 AYNRMPLPIVMSLTVLIGITLFP-PPESLGMTPLE 517  
Db 429 AY-----VFIVTVFLVIFWVTFEKKVPETGRTFEE 460

RESULT 11  
US-08-355-844-3  
; Sequence 3, Application US/08355844  
; Patent No. 5940307  
; GENERAL INFORMATION:  
; APPLICANT: Fischbarg, Jorge  
; APPLICANT: Czegledy, Ferenc  
; APPLICANT: Iserovich, Pavel  
; APPLICANT: Li, Jun  
; APPLICANT: Cheung, Min  
; TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN  
; TITLE OF INVENTION: STRUCTURE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/355,844  
; FILING DATE: 14-DEC-1994  
; CLASSIFICATION: 436  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tang, Henry Y.S.  
; REGISTRATION NUMBER: 29,705  
; REFERENCE/DOCKET NUMBER: A29927-50/29910  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-408-2586  
; TELEFAX: 212-765-2519  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 492 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: 1..492  
; OTHER INFORMATION: Facilitative glucose transportor  
; OTHER INFORMATION: Glut1 protein  
US-08-355-844-3

Query Match  
Best Local Similarity 7.8%; Score 222; DB 2; Length 492;  
Matches 117; Conservative 74; Mismatches 175; Indels 122; Gaps 21;

QY 119 FSDQVYLSTVVTENMLVCEENKWKVPLTSLFFVGVLLGVSFVSGQLSDRGRK-----NV 172  
DB 52 YGESLPTLTTLTWS-----LSVAIFSVMGMSFVSGLFVFNRRGRNSMLMNL 101  
QY 173 LFATMAVOTGFSFLQIFSIWEM-----FTVLVIVGNGOISNVVAFILGT 219  
DB 102 LAFVSAVLGMSKLGKSEFEMLLGRFIIGVYCGLTTCFVPMYVGEVSPFAFRGALGT 158  
QY 220 EILGKSVRIEFTLGCTFFAVGYMLPLPAYFI-----RD-WRMLLALTVPGLCVPLM 274  
DB 159 -----LHQLGI-----VVGILIAQVFGDLSIMGKDLWPLLSIIFIPALLQCI 204  
QY 275 WFIPESPRWLISQRRFREAEDIIQAKAMNNTAVPAVIFDSVEELNPLKQ-----K 326  
DB 205 PCPESPRELLINNEENRAKSVLK--KLGRGA-----DVTHDLOEMKEESQMMREKK 256  
QY 327 AFILDLFTRN-----IAIMTMSLLMLTSGVGFALSADPNLHGDAYLNCFLSALI 380  
DB 257 VTILEFRSPAYRQPIILIAVVLQSLQSGINAVFYTSI-----FEKAGV 303  
QY 381 EIPAVIT-----AWLLRLTPRR--YIIAAVLFWGGVLLFIQLVPVDYV-- 423  
DB 304 QQPVTATIGSGIVNTAFTVSVLFVVRAGRRTLHLGLAGMAGQAIIIMTIALLEQLPW 363  
QY 424 --FLSIGLVMGKFGITSAFS-----MLYVFTAELYPTLVNMAVGVSTASRVGS-ITA 475  
DB 364 MSYLSIVAI-----FGVAFVGVGPPIPFVIAELESQGPRAAIAVAGFSNWTNFI 419  
QY 476 PFYVYLGAYNRMPLVYVNGSLTVLIGITLFPESIGMTLPETLQMKVKWFRSGKKT 534  
DB 420 MCFQVVEQLCGPYVEIIFTVLLVFFIRTYKVPETAGRTFDEI-----ASGFRQGGAS 473  
QY 535 RDSMETEE 542  
DB 474 QSDKTPEE 481

## RESULT 12

PCT-US95-16126-3

Sequence 3, Application PC/TUS9516126

## GENERAL INFORMATION:

APPLICANT: Fischbarg, Jorge  
APPLICANT: Czegledy, Ferenc  
APPLICANT: Iserovich, Pavel  
APPLICANT: Li, Jun  
APPLICANT: Cheung, Min  
TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN  
STRUCTURE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10112-0228

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16126  
FILING DATE:

## CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/355,844

FILING DATE: 14-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Tang, Henry Y.S.  
REGISTRATION NUMBER: 29,705  
REFERENCE/DOCKET NUMBER: A29927-50/29910  
TELEPHONE: 212-408-2586  
TELEFAX: 212-765-2519  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 492 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Human  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..492  
OTHER INFORMATION: Facilitative glucose transporter  
OTHER INFORMATION: Glut1 protein  
PCT-US95-16126-3

Query Match 7.8%; Score 222; DB 5; Length 492;

Best Local Similarity 24.0%; Pred. No. 9.7e-15;

Matches 117; Conservative 74; Mismatches 175; Indels 122; Gaps 21;

QY 119 FSDQVYLSTVVTENMLVCEENKWKVPLTSLFFVGVLLGVSFVSGQLSDRGRK-----NV 172  
DB 52 YGESLPTLTTLTWS-----LSVAIFSVMGMSFVSGLFVFNRRGRNSMLMNL 101  
QY 173 LFATMAVOTGFSFLQIFSIWEM-----FTVLVIVGNGOISNVVAFILGT 219  
DB 102 LAFVSAVLGMSKLGKSEFEMLLGRFIIGVYCGLTTCFVPMYVGEVSPFAFRGALGT 158  
QY 220 EILGKSVRIEFTLGCTFFAVGYMLPLPAYFI-----RD-WRMLLALTVPGLCVPLM 274  
DB 159 -----LHQLGI-----VVGILIAQVFGDLSIMGKDLWPLLSIIFIPALLQCI 204  
QY 275 WFIPESPRWLISQRRFREAEDIIQAKAMNNTAVPAVIFDSVEELNPLKQ-----K 326  
DB 205 PCPESPRELLINNEENRAKSVLK--KLGRGA-----DVTHDLOEMKEESQMMREKK 256  
QY 327 AFILDLFTRN-----IAIMTMSLLMLTSGVGFALSADPNLHGDAYLNCFLSALI 380  
DB 257 VTILEFRSPAYRQPIILIAVVLQSLQSGINAVFYTSI-----FEKAGV 303  
QY 381 EIPAVIT-----AWLLRLTPRR--YIIAAVLFWGGVLLFIQLVPVDYV-- 423  
DB 304 QQPVTATIGSGIVNTAFTVSVLFVVRAGRRTLHLGLAGMAGQAIIIMTIALLEQLPW 363  
QY 424 --FLSIGLVMGKFGITSAFS-----MLYVFTAELYPTLVNMAVGVSTASRVGS-ITA 475  
DB 364 MSYLSIVAI-----FGVAFVGVGPPIPFVIAELESQGPRAAIAVAGFSNWTNFI 419  
QY 476 PFYVYLGAYNRMPLVYVNGSLTVLIGITLFPESIGMTLPETLQMKVKWFRSGKKT 534  
DB 420 MCFQVVEQLCGPYVEIIFTVLLVFFIRTYKVPETAGRTFDEI-----ASGFRQGGAS 473  
QY 535 RDSMETEE 542  
DB 474 QSDKTPEE 481

## RESULT 13

US-09-031-392-10

Sequence 10, Application US/09031392

Patent No. 5942398

## GENERAL INFORMATION:

APPLICANT: Tartaglia, Louis A.

APPLICANT: Weng, Xun

TITLE OF INVENTION: NUCLEIC ACID MOLECULES

TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF

Fri Jan 25 10:19:00 2002

NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/031,392  
FILING DATE: 26-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Meiklejohn, Ph.D., Anita L.  
REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 07334/072001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 493 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-031-392-10

Patent No. 6136547  
GENERAL INFORMATION:  
APPLICANT: Tartaglia, Louis A.  
ADDRESSEE: Weng, Xun  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
GLUTEX AND USES THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/299,549  
FILING DATE: 26-APR-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/031,392  
FILING DATE: 26-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Meiklejohn, Ph.D., Anita L.  
REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 07334/072002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 493 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-299-549-10

Query Match 7.6%; Score 216; DB 2; Length 493;  
Best Local Similarity 24.2%; Pred. No. 4.le-14;  
Matches 111; Conservative 78; Mismatches 171; Indels 98; Gaps 19;

QY 144 LTTSLFFVGLGSGVSGQLSDRGKKNVLEA--TMAVOTGFSEF-LQIFSIEMFTVLF 200  
DB 68 LSVSIFAVGGMIGSLVXGNLGRKXAMLVNNVLAAGLLMGLAKXAXSEFMILGR 127

QY 201 VIVGM-----GOISNVYVAFILGT-EILGKSVRIIFS-TLGVCTFFAVGYML 245  
DB 128 FIILGCLSSGVVPMVYGEISPTALRGALGTNLQIGVIGILIAQVLGLDSL--LGNES 185

QY 246 LPLFAYFIRDMRLALLTPGVLCVPLWFWFIPESRWLI-----SORREREAE 294  
DB 186 L-----WPLLGLTGPALLQLLPPCPSPRYLLINKNEEARAKKALORLRTA 236

QY 295 DIQKAARMNTAVPAVIFDSVEELNPLKQKAFILDLFRTN-----IAIMTMSLLL 348  
DB 237 DVSQEAEMKD-----ESRXMXSEKXSVLELFRSRXYRQPVIIAIVLQLSQOL 285

QY 349 WMLTSVGYFALSADPNLHGDAYLNCFLSALIEIPAYIT-----AWLLLRTL 395  
DB 286 SGINAVFYTSTI-----FEKAGVQPVYATIGAGVNVNFTVVSFVVERA 332

QY 396 PRP--YIIAAVLFWGGVLLFIQLVPVDY--YFLSIGVLMGKFGITSASF-----MLYV 446  
DB 333 GRRTLHLGLGGMAGCAVLMTIALLDQVPWMSYVSIVAI--FGFVAFEEVGPPIPW 390

QY 447 FTAELYPTLVRNMAVGVTSTASRVGS--IIAPYFVYLGAYNRMPIYVMSGLSVLIGIFT- 504  
DB 391 IVAELFSQGRPPAAIAVAGFNSNWTNFIIVGLLFQYIAELLGPYFIVFAVLLLFFITFF 450

RESULT 14

US-09-299-549-10

; Sequence 10, Application US/09299549







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 10:08:24 ; Search time 28.31 seconds  
(without alignments)  
1482.592 Million cell updates/sec

Title: US-09-521-195-1  
Perfect score: 2845  
Sequence: 1 MRDYDEVIAFLGEGFPQRL.....KKTRDSMETEENPKVLITAF 551

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2218	78.0	557	2 JW0089	organic cation tra
2	2152	75.6	557	2 JE0346	high-affinity carn
3	754.5	26.5	576	2 T22509	hypothetical prote
4	699.5	24.6	593	2 JC4884	organic cation tra
5	681.5	24.0	556	2 S50862	organic cation tra
6	565	19.9	794	2 T27870	hypothetical prote
7	548	19.3	745	2 T16565	hypothetical prote
8	484.5	17.0	527	2 T01019	transport protein
9	475	16.7	539	2 C96758	probable protein
10	421.5	14.8	518	2 B86299	hypothetical prote
11	412.5	14.5	521	2 H86298	hypothetical prote
12	382.5	13.4	515	2 B96825	hypothetical prote
13	356	12.5	528	2 T21682	hypothetical prote
14	338	11.9	540	2 T25851	hypothetical prote
15	327	11.5	454	2 F73580	probable sugar tra
16	317	11.1	455	2 B83213	probable MPS trans
17	311	10.9	591	2 T30895	sugar transport pr
18	307	10.8	435	2 T15290	hypothetical prote
19	294.5	10.4	529	2 T23190	hypothetical prote
20	290	10.2	400	2 C69757	transporter homolo
21	286.5	10.1	478	2 T33985	hypothetical prote
22	281	9.9	422	2 G72234	hypothetical prote
23	276	9.7	461	2 D70073	metabolite transpo
24	275	9.7	520	2 T23545	hypothetical prote
25	274.5	9.6	443	2 E64725	yaaU protein - Esc
26	274.5	9.6	443	2 H85485	probable transport
27	272.5	9.6	442	2 A83122	probable MPS trans
28	271.5	9.5	524	2 T27082	hypothetical prote
29	271	9.5	423	2 T19030	hypothetical prote

ALIGNMENTS

RESULT 1

JW0089  
organic cation transporter protein 2 - human  
N;Alternate names: OCTN2  
C;Species: Homo sapiens (man)  
C;Date: 18-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 28-May-1999  
C;Accession: JW0089  
R;Wu, X.; Prasad, P.D.; Leibach, F.H.; Ganapathy, V.  
Biochem. Biophys. Res. Commun. 246, 589-595, 1998  
A;Title: cDNA sequence, transport function, and genomic organization of human OCTN2,  
A;Reference number: JW0089; MUID:98289574  
A;Accession: JW0089  
A;Molecule type: mRNA  
A;Residues: 1-557 <WUA>  
A;Cross-references: GB:AF057164; MID:g3273740; PIDN:AAC24828.1; PID:g3273741  
A;Experimental source: placenta  
C;Comment: This transporter functions in the elimination of cationic drugs and other

Query Match 78.0%; Score 2218; DB 2: Length 557;  
Best Local Similarity 75.9%; Pred. No. 1.6e-156;  
Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

Qy	1	MRDYDEVIAFLGEGFPFORLIFFLLSASIIIPNGFMGSSVFLAGTPEHRCRVPDAANLSS	60
Db	1	MRDYDEVIAFLGEGFPFORLIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS	60
Qy	61	AWRNNVPLRLRDGREVPHSCRYRLATIANFSAIGLEPGRDVDLQLEQESCLDGWEFS	120
Db	61	AWRNNVPLRLRDGREVPHSCRYRLATIANFSAIGLEPGRDVDLQLEQESCLDGWEFS	120
Qy	121	QDVYLSTVVTWNVLVCEDNWKVPLTTSLFFVGLLGSFVSGQLSDRFGKKNVLEATMAVQ	180
Db	121	QDVYLSTVVTWNVLVCEDNWKVPLTTSLFFVGLLGSFVSGQLSDRFGKKNVLEATMAVQ	180
Qy	181	TGFSFLQIFSIQSWEMFTLVFVIVGMGQISNVVAFILGTILGKSVRIIFSTGLGVCFFFA	240
Db	181	TGFSFLQIFSIQSWEMFTLVFVIVGMGQISNVVAFILGTILGKSVRIIFSTGLGVCFFFA	240
Qy	241	VGYMLLPLFAFIRDRWMLLALTPGVCLVPLWFWFIPESPRWLLISQRRFREAEDIQKA	300
Db	241	VGYMLLPLFAFIRDRWMLLALTPGVCLVPLWFWFIPESPRWLLISQRRFREAEDIQKA	300
Qy	301	AKMNNTPAVTFD--SVEELNPLKQKAFILDLPRTNIAIMTISMLLWMLTSVGYFA	358
Db	301	AKANGIVPSTIFDSELDQSSKKQSHNILLRTWNIRMTIMSLMWTLSVGYFG	360
Qy	359	LSLDAPNLHGDAYLNCFLSALIEIPAYITAMLLLTLPRLRYIIAAVLPWGGLVLTQLV	418
Db	361	LSLDTPNLHGDIYFNCFLSAMVEVPAYVLAWLLQLYLPRLRYMATATFLGGSVLLFMOLV	420
Qy	419	PVDYFSLTGLVLMKFGITSAFSLYVFTAEYFTPLVRNMAVGVTSTASRGSIIAPYF	478

probable sugar tra  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hexose transport p  
synaptic vesicle p  
hypothetical prote  
glucose transport  
hexose transport p  
glucose transport  
probable transport  
glucose transport  
hypothetical prote  
transmembrane tran  
glucose transport  
glucose transport

Db 421 PPDIYLAIVLVYSGKGVTAFAFVNVYVYTAELVPTVVRNMGVSVSTASRLGSLSPYF 480  
 Qy 479 VYLGAYNRMPLVYIVVWGSITVLIGITLFFPFESLGMTLPTLEQMKVKWFRSGK---KTR 535  
 Db 481 VYLGAYNRMPLVYIVVWGSITVLIGITLFFPFESLGMTLPTLEQMKVKWFRSGK---KTR 540  
 Qy 536 DSMETEENPKVL-ITAF 551  
 Db 541 MKDQGERPTILKSTAF 557

RESULT 2  
 JE0346  
 high-affinity carnitine transporter, CT1 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
 C:Accession: JE0346  
 R:Sekine, T.; Kusuhara, H.; Utsunomiya-Tate, N.; Tsuda, M.; Sugiyama, Y.; Kanai, Y.; Endo  
 Biochem. Biophys. Res. Commun. 251, 586-591, 1998  
 A:Title: Molecular cloning and characterization of high-affinity carnitine transporter  
 A:Reference number: JE0346; MUID:99011422  
 A:Accession: JE0346  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-557 <SEK>  
 A:Cross-references: DDBJ:AB017260; NID:g3869208; PIDN:BAA34399.1; PID:g3869209

Query Match 75.6%; Score 2152; DB 2; Length 557;  
 Best Local Similarity 73.1%; Pred. No. 1.2e-151;  
 Matches 407; Conservative 68; Mismatches 76; Indels 6; Gaps 3;

Qy 1 MRDYDEVIATFLGEMGPQRLIFLLS-ASIIIPNGFNMGSVVFLAGTPEHRCRVPDAAVLSS 60  
 Db 1 MRDYDEVIATFLGEMGPQRLIFLLS-ASIIIPNGFNMGSVVFLAGTPEHRCRVPDAAVLSS 60  
 Qy 61 AWRNNSVPLRLDGRVPHSCRYRLATIANFSALGLEPGRVDLQLEQESCLDGEWFS 120  
 Db 61 AWRNNSVPLRLDGRVPHSCRYRLATIANFSALGLEPGRVDLQLEQESCLDGEWFS 120  
 Qy 121 QDYVLSVTVTEWNLVCEENKWKVPLTSLFFVGVLLGSFVSGQLSDFRGKKNVLFATMAVQ 180  
 Db 121 KDVFSLTIVTEWDLCKDKWKAPLITSLFFVGVLLGSFVSGQLSDFRGKKNVLFATMAVQ 180  
 Qy 181 TGSFLOIFSISWEMFTVLVIVGMQISNYVVAEILGTILGKSVRIIFSTLGVCTFFA 240  
 Db 181 TGSFLOIFSISWEMFTVLVIVGMQISNYVVAEILGTILGKSVRIIFSTLGVCTFFA 240  
 Qy 241 VGYMLLPFAFYTRDWRMLLALTVPGLVPLWVFIPESPRWLISQRRFREAEDIIQKA 300  
 Db 241 VGYMLLPFAFYTRDWRMLLALTVPGLVPLWVFIPESPRWLISQRRFREAEDIIQKA 300  
 Qy 301 AKNNNTAVPAVIFD--SVBELNPLKQKAFILDLFRTRNIAIMTMSLLMLTMSVGYFA 358  
 Db 301 AKNNNTAVPAVIFD--SVBELNPLKQKAFILDLFRTRNIAIMTMSLLMLTMSVGYFA 358  
 Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYTIAAVLFWGGVLLFIQV 418  
 Db 361 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYTIAAVLFWGGVLLFIQV 420  
 Qy 419 PVDYVFLSISGLVNLKFGTSAFSLMYVFTAEYPTLVLRNMAVGTSTASRVGSI 478  
 Db 421 PSELVFLSTALVWVGKFGTSAFSLMYVFTAEYPTLVLRNMAVGTSTASRVGSI 480  
 Qy 479 VYLGAYNRMPLVYIVVWGSITVLIGITLFFPFESLGMTLPTLEQMKVKWFRSGK---KTR 535  
 Db 481 VYLGAYNRMPLVYIVVWGSITVLIGITLFFPFESLGMTLPTLEQMKVKWFRSGK---KTR 540  
 Qy 536 DSMETEENPKVL-ITAF 551  
 Db 541 TOKDGGESPTVLKSTAF 557

RESULT 3  
 T22509  
 hypothetical protein F52F12.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T22509  
 R:Matthews, L.  
 submitted to the EMBL Data Library, December 1996  
 A:Reference number: Z19573  
 A:Accession: T22509  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-576 <WIL>  
 A:Cross-references: EMBL:Z83228; PIDN:CAB05732.1; GSPDB:GN00019; CESP:F52F12.1  
 A:Experimental source: clone F52F12  
 C:Genetics:  
 A:Gene: CESP:F52F12.1  
 A:Map position: 1  
 A:Introns: 16/1; 23/3; 50/1; 80/3; 108/2; 134/3; 221/1; 257/2; 319/3; 358/2; 490/3; 5

Query Match 26.5%; Score 754.5; DB 2; Length 576;  
 Best Local Similarity 30.2%; Pred. No. 2.8e-48;  
 Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;

Qy 3 DYDEVIAFLGEMGPQRLIFLLS-ASIIIPNGFNMGSVVFLAGTPEHRCRVPDAAVLSS 61  
 Db 16 DFDVLEQVGNVYQIVFFFIICLPTSLPSAFSAFNIPFVVGNNPHTCHIPEKGEYLRP 75  
 Qy 62 WRNNSVPLRLDGRVPHSCRYRLATIANFSALGLEPGRVDLQLEQESCLDGEWFSQ 121  
 Db 76 LTNDQFILL-----SCHQYNETQINVFRAFTSAP-VDTYSDRISLVPCQNGWDIDN 124  
 Qy 122 DVLSTVTVTEWNLVCEENKWKVPLTSLFFVGVLLGSFVSGQLSDFRGKKNVLFATMAVQ 181  
 Db 125 STYLDLSLVEENLVCDOQAWIEITSTTFYVGSFIGNCLFGYVADKFGRRSRFFVILTVLI 184  
 Qy 182 GFSFLOIFSISWEMFTVLVIVGMQISNYVVAEILGTILGKSVRIIFSTLGVCTFFAV 241  
 Db 185 VGTASSFAKDIETFIILURFTGLAFPAFQIPFIICMEFMGNSGR-IFSGLMTSLFFGA 243  
 Qy 242 GYMLLPFAFYTRDWRMLLALTVPGLVPLWVFIPESPRWLISQRRFREAEDIIQKA 301  
 Db 244 AMALLGVVAMFIRWRQLTFFCNAPAFYIIYFELPESPRWSVSGKADAKKQLKKA 303  
 Qy 302 KMN--NTAVPAVIFDSVEELNPLKQKAF-----ILDLFRTRNIAIMTMSLLMLT 353  
 Db 304 KMNKSNVDVDELV-DSMKNHQNAEEKETKRSHNVTDLEKTPNLRKTLIVTYIWMNA 362  
 Qy 354 VGYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYTIAAVLFWGGVLL 413  
 Db 363 IYVGLNLTNLSNLPVDDYWSFIINGAVELPGYFVWPLLCACGRMTTAAATMIVCGICV 422  
 Qy 414 FIQLVVDYVFLSISGLVNLKFGTSAFSLMYVFTAEYPTLVLRNMAVGTSTASRVGSI 473  
 Db 423 SAMFMPDGPWLVASAFSGFKGVSGFAVIYIFAGELYPTVVRRAIGMGSSNVASGGL 482  
 Qy 474 IAPVYVYLGAYNRMPLVYIVVWGSITVLIGITLFFPFESLGMTLPTLEQMKVKWFRSGK 533  
 Db 483 LAPHIVNLGKIVLILLPLMLLMALSAGILTFPLPETLGLAPLMTTIEDAENF-----GKK 537  
 Qy 534 -----TRDSMETEENP 544  
 Db 538 PEPDSGMFTQAACKRESQP 556

RESULT 4  
 JC4884  
 organic cation transporter protein 2 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 10-Sep-1996 #sequence\_revision 18-Oct-1996 #text\_change 05-Nov-1999  
 C:Accession: JC4884



R:Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.  
 Biochem. Biophys. Res. Commun. 224, 500-507, 1996  
 A:Title: cDNA cloning and functional expression of a novel rat kidney organic cation tra  
 A:Reference number: JC4884; MUID:96295517  
 A:Accession: JC4884  
 A:Molecule type: mRNA  
 A:Residues: 1-593 <ORU>  
 A:Cross-references: DBJ:D83044; NID:g1502282; PIDN:BAAL1754.1; PID:g15022  
 A:Experimental source: kidney  
 C:Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 24.6%; Score 699.5; DB 2; Length 593;  
 Best Local Similarity 34.0%; Pred. No. 3.4e-44;  
 Matches 189; Conservative 87; Mismatches 243; Indels 37; Gaps 14;  
 QY 1 MRDYDEVATFAGCGPQFORLIFFLLSASLIIPNGFNG--MSVVFAGTPEHRCRVPDAANL 58  
 DB 1 MSTVDDILEHGEFHLFOKQTFLL--ALLSGATPIYGVIVLGFTHCHWSPGAKL 58  
 QY 59 SS--AWR-----NNSVP--LRLRDGREVPHSCSYR-----LATIANFSALGLEPGRDV 103  
 DB 59 SQRCSQAEEELNYTPVGLSPDEASFLSCMRVEVDNQSTLDCVDPPLSLA-----A 112  
 QY 104 DLGQLEQSCLDGWFSDQVYLSTVVTENLVCEENKVPVLTSLFFVGLLGSFVSQOL 163  
 DB 113 DRNQLPLFCERHGWYNTFG--SSIVTEFNLVCAHSMMLDFQSVVNVGFFIGAMIGYL 170  
 QY 164 SDRFGKKNVLFATMAVOTGFSFLQIFSIWEMFTVLFVIVGMGQISNVYVAFILGTEILG 223  
 DB 171 ADRFGKCLLVITILINISGALMAISPAYAWMLVFRFLOGLVSKAGWLGIVILITEFVG 230  
 QY 224 KSVRIIFSTLGYC--TFPAGVMYLLPLFAFYFIRDRWMLLALLTPGVLCVPLWFIPEPSP 281  
 DB 231 LGYR---RMVGICYQIAFTVGLLILAGVAYIPNNRWLQFAVTLFNFCLLYFMCIPESP 287  
 QY 282 RWLISORFREADIIOAKKNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNRNIAIM 341  
 DB 288 RWLISONKIVKAMIKIHKAKNGKSVPSVLQNLTPEDAGKKLPSILDLVTPQIRKH 347  
 QY 342 TMSLLMLMTLVGVFALSALDAPNLHGD--AYLNCFLSALIEPAITAWLLRLTPRYI 400  
 DB 348 TLILMYNFTSSVLYQGLIMHM--GLAGDNIYLDFFYSALVEPFAAFIILITIDRVGRYP 406  
 QY 401 IAAVLFWGGVLLFIQVPDYFYSIGLVMLKFGITSAFSLVYVFAELVPTLVRNA 460  
 DB 407 WAYSNNVAGAACLASVFIIDDLQWLKITIACLRMGITMAYEMVCLVNAELVPTIIRNLG 466  
 QY 461 VGTSTASRVGSIAPFYV--LGAYNRMLPYIVMGSLTVLIGITLFFPESLGMTLPETL 519  
 DB 467 VLVCSMCDIGIITFLVYRLTDIWMFPLVFAVVGVLVAGALVLLPETKKGALPETI 526  
 QY 520 EQMKVWFRSGKKTR 535  
 DB 527 EDAAENMQ--RPRKKER 540

RESULT 5  
 S50862  
 Organic cation transport protein OCT1 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 05-Nov-1999  
 C:Accession: S50862; S78533; I58089  
 R:Gruendemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.  
 Nature 372, 549-552, 1994  
 A:Title: Drug excretion mediated by a new prototype of polyspecific transporter.  
 A:Reference number: I58089; MUID:95082907  
 A:Accession: S50862  
 A:Molecule type: mRNA  
 A:Residues: 1-556 <ORU>  
 A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622  
 A:Experimental source: kidney  
 R:Gorboulev, V.G.

submitted to the EMBL Data Library, January 1995

A:Reference number: S78533  
 A:Accession: S78533  
 A:Molecule type: mRNA  
 A:Residues: 1-342, 'N', 344-556 <OR>  
 A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622  
 C:Keywords: glycoprotein; phosphoprotein; transmembrane protein  
 F:20-46/Domain: transmembrane #status predicted <TM1>  
 F:154-171/Domain: transmembrane #status predicted <TM2>  
 F:178-197/Domain: transmembrane #status predicted <TM3>  
 F:243-260/Domain: transmembrane #status predicted <TM4>  
 F:267-283/Domain: transmembrane #status predicted <TM5>  
 F:350-366/Domain: transmembrane #status predicted <TM6>  
 F:380-398/Domain: transmembrane #status predicted <TM7>  
 F:406-425/Domain: transmembrane #status predicted <TM8>  
 F:435-452/Domain: transmembrane #status predicted <TM9>  
 F:469-485/Domain: transmembrane #status predicted <TM10>  
 F:494-514/Domain: transmembrane #status predicted <TM11>  
 F:71,97,113,432/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:286,292/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pred  
 F:296,343,550/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status

Query Match 24.0%; Score 681.5; DB 2; Length 556;  
 Best Local Similarity 33.0%; Pred. No. 6.8e-43;  
 Matches 183; Conservative 88; Mismatches 231; Indels 53; Gaps 11;  
 QY 1 MRDYDEVATFAGCGPQFORLIFFLL--SASLIIPNGFNGMSVVFAGTPEHRCRVPDAAN 57  
 DB 1 MPTVDDVLEQGEFGFQKQAFLLCLLISASLAP---IYGVIVLGFTHCHWSPGVAE 57  
 QY 58 LSS--AWR-----NNSVP--LRLRDGREVPHSCSYR-----ATIANFSAL 95  
 DB 58 LSQRCSQAEEELNYTPVGLSPDEASFLSCMRVEVDNQSTLDCVDPPLSLVANSRQL 117  
 QY 96 GLEPRDVLGQLEQSCLDGWFSDQVYLSTVVTENLVCEENKVPVLTSLFFVGVLL 155  
 DB 118 PLGP-----CEHGWY--DTPGSSIVTEFNLVCGDAWKVDLFQSCVNLGFFL 162  
 QY 156 GSFVSGQLSDRFGRKNVLFATMAVOTGFSFLQIFSIWEMFTVLFVIVGMGQISNVYVAF 215  
 DB 163 GSVVGVYADRFGRKCLLVLTLYTSVSGVLTAVAPDYTSMLLFRLLQGMVSKGWSGY 222  
 QY 216 ILGTEILGKSVR-----IFSTLGVCTFFAGVMYLLPLFAFYFIRDRWMLLALLTPGVLCV 271  
 DB 223 TLITEFVSGYRRTAILYQMA-----FTVLVGLAGVAYAPDWRWLQAVSLPTLFL 277  
 QY 272 PLWFIPESPRLISORFREADIIOAKKNNTAVPAVIFDSVEELNPLKQKAFIL 331  
 DB 278 LYWFVPESPRLISQKRTTRAVRIMEQIAOKNGKVPADLKMCLCEEDASEKRSFAD 337  
 QY 332 LFRTRNIAIMTMSLLMLTSGVFALSALDAPNLHGDAYLNCFLSALIEPAITAWLL 391  
 DB 338 LFRPTLRKHVILMYLWFSCAVLYQGLIMHVGATGANLYLDFFYSALVEPFAAFIILVT 397  
 QY 392 LRTPLPRVYIAAVLFWGGVLLFIQVPDYFYSIGLVMLKFGITSAFSLVYVFAEL 451  
 DB 398 IDRIGRIYPIAASNLVTGAACLLMIFIPHEHLNLTACLRMGATIVLQMVCLVNAEL 457  
 QY 452 YPTLVNRNAVGTSPASRVGSIAPFYV--LGAYNRMLPYIVMGSLTVLIGITLFFPES 510  
 DB 458 YPTFIRNLGMVCSALCDLGGITFPFWFRLMEVWQALPLILFGVLGTAGAMTLLPET 517  
 QY 511 LGMTLPLEQMKV 525  
 DB 518 KGVALPETIEEAENL 532

RESULT 6  
 T27870  
 hypothetical protein ZK455.8 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C;Accession: T27870

R;White, S.

submitted to the EMBL Data Library, November 1995

A;Reference number: Z20432

A;Accession: T27870

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-794 <WIL>

A;Cross-references: EMBL:Z66567; PIDN:CAA91492.2; GSPDB:GN00028; CESP:ZK455.8

A;Experimental source: clone ZK455

C;Genetics:

A;Gene: CESP:ZK455.8

A;Map position: X

A;Introns: 52/3; 119/2; 158/1; 207/3; 291/3; 335/2; 428/3; 610/2; 636/3; 733/3

Query Match 19.9% Score 565; DB 2: Length 794;

Best Local Similarity 26.4% Pred No. 4.2e-34;

Matches 158; Conservative 121; Mismatches 232; Indels 88; Gaps 20;

Qy 3 DYDE-----VIAFLGPGPQRLIFLLSASIIIPNGFNGMSVVFVLAGTPEHRCRVP-----53

Db 155 DFEEDLILQIGCSYQWIIIVLIISVQVQPHAMFNLSVVMYQPDHWCXIPFNEE 214

Qy 54 ----DAANLSSAW--RNNSV--PLRLRDGVEPH--SCSRRLATTANFSAALGLEPGRD 102

Db 215 SFSAEGLTNTWQVLSNTIAPRTENKORNELHDOCHYE----RDYVHKLSFWAQ 270

Qy 103 V-DLQQLQESCLDQWESQDYLSTVTEWNLVCEDNWKVPLTTLSTLFFVGLLSFVSG 161

Db 271 VKDMNATKINRCKEWEYDTSMDRTIVTEWNRVCDNNSRAHVMSYSLGYLVGCFVGG 330

Qy 162 OLSDRFGRKKNVLFATMAVQTGFSFLOIFSISWEMFTVL-FVIIVGMGOISNYVAFILGTE 220

Db 331 FIDRIYGRKTAITGEGILTMFLGELLYSKEPELFLVRELLAATNEAAD-LAAVYICWE 389

Qy 221 ILGKSVRIIFSTLGVCTFFFAVGYMLLFLFYFIRDMRLMLLALTLPVGLCVPLWMPFIPES 280

Db 390 VTGTYRSIVGSL-IOAPWAGYAFALAIYATKSTMTMHLICVLLHIIISMLLYPLPS 448

Qy 281 PRLWLSORREAEEDIIOKAKMNTAPPA--VIFDSVEELNPLK-QOKAFILDLPRTN 337

Db 449 PRLWLNKTKOAEKILREACHYKNRSLPSDLGLVLRHAERKKWKWKUNERKPSYFLHFRSE 508

Qy 338 TAIMTISMLLWM-----LTSVGY-----356

Db 509 LFRNRVLFIVWVTKNIFONSREFOYSDNSCGLLRNGYCVIRSVITRKVGRTLSEKFR 568

Qy 357 -----FALSIDAPNLKGDAYLNCFLSALIELPAYITAWLLRLPLPRYIIAALVFWGGY 411

Db 569 NOKLCFRVFD-----GNFFLNAMAGAIELPTLVFCVFLLR-MGRKRSQMLVLF-GSGL 621

Qy 412 LLFTQLVPV--DYFLSLGLVMLGKFGITSAFSLMYVFTAEYPTLVRNMAVGTSTASR 469

Db 622 FLTTSVVMYVRKQSTLALIFLMLSKACIQGSFNILYFTSELNPTVVRNSAVGSISSMAR 681

Qy 470 VGSIIAPFVVLGAYNR-MLPYIVMGSLTVLIGITLFFPESLGMTLPET-LEOMQVKV 526

Db 682 MGAGASGIIALSDVWPLVPMWTFACFSLACGLVLLLPETOGCLPDPDIIILDSVQNR 740

RESULT 7

Ti6565

hypothetical protein K05F1.6 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C;Accession: Ti6565

R;Wohlmann, P.

submitted to the EMBL Data Library, June 1995

A;Description: The sequence of C. elegans cosmid K05F1.

A;Reference number: Z1853

A;Accession: Ti6565

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-745 <WOH>

A;Cross-references: EMBL:U29377; NID:g868173; PID:g868176; PIDN:AAA68713.1; CESP:K05F

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:K05F1.6

A;Introns: 70/2; 160/3; 205/3; 249/3; 293/2; 350/1; 386/3; 462/3; 546/2; 669/3; 680/3

Query Match 19.3% Score 548; DB 2: Length 745;

Best Local Similarity 27.8% Pred. No. 7e-33;

Matches 166; Conservative 112; Mismatches 233; Indels 86; Gaps 21;

Qy 3 DYDEVIAFLGEPGQRLIFLLSASIIIPNGFNGMSVVFVLAGTPEHRCRVP-----53

Db 112 DFEILNIIGCRWQIWIIVLIALQIPHAMFNLSVVMYQPDHWCXIPFNEE 171

Qy 54 -----DAANLSSAWRN-NSVPLRLRDGVEPHSC-----SRYLATIANFS 93

Db 172 ISSHGPLYMGVDDIKNISVFPNANS-----DGAVQDSCFYERSEERYKQLRMPLE 225

Qy 94 ALGLEPGRDVLGQLEQESCLDQWESQDYLSTVTEWNLVCEDNWKVPLTTLSTLFF-VG 152

Db 226 TAMSEANKDV-----APKKKC-QAYHFEKDVVETIVTDENLVC-DSWFAKGAHMFYSIG 279

Qy 153 VLLGSFVSGQLSDRFGRKKNVLFATMAVQTGFSFLOIFSISWEMFTVL-FVIIVGMGOISNYV 212

Db 280 YLLCGVGGTASDKIGRKPTIIIGFGLSSMLGVLPFDNYPMFLRLILSLAICNEAADL 339

Qy 213 VAFILGTEILGKSVRIIFSTLGVCTFFFAVGYMLLFLFYFIRDMRLMLLALTLPVGLCV 272

Db 340 AAYTLCMEITGTYRAMVGSMLQAPWALGYALLAIYATKSTMTMHLICVLLHIIISMLLY 398

Qy 273 LWNIPSPRWLISORREAEEDIIOKAK-----MNNTA-----VPA---VIFDSVEE 318

Db 399 FICSPSPRWLMVQNVSEAEVIRKACREPPFPFNMCCTTKCGNLPSDLDELIVSHREK 458

Qy 319 LNPILKQKRAFILDTRTNIAIMTISLLMLTSLVGYFALS-----DAPN---IHGDA 370

Db 459 LNKNGKIGFLDLFTMKELRYRTISVCIVFEMATALVYVGLVMAVSDQAPGRTLTGYF 518

Qy 371 YLNCFLSALTEIPA-YITAWLLRLPLPRYIIAALVFWGGVLLFTQLVPVD-YFELSIG 428

Db 519 HLNGIAGATIEPIPLFACVWMM--QLGRKALMLTLITSGLFTIIVAMLSVSGHYMLALA 576

Qy 429 LVMLGKFGITSAFSLMYVFTAEYPTLVRNMAVGTSTASRVGSIIAPYFVILGAYNRL 488

Db 577 FMYGKIAGVQAFNLIYFTSELYPTVVRNTAVGTVSMVAREGSLSYIALLS--NISL 634

Qy 489 PYIVMGSLTVLIGITLTF-----FPESLGMWLPETLEO-----MOKVWFRSGKKT 534

Db 635 PIVPM---IFAVESLFAGLMVLVLPETSEKPLPETDDDAINFLEPTKQFTIEST 687

RESULT 8

T01019

transport protein homolog YUP8H12R.2 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 22-Oct-1999

C;Accession: T01019

R;Theologis, A.; Vysotskaya, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K

Definer, P.; Davis, R.W.

submitted to the EMBL Data Library, May 1998

A;Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.

A;Reference number: Z14227

A;Accession: T01019

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-527 <THE>

A;Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152576; GSPDB:GN00059; ATSP:YU

A;Experimental source: cultivar Columbia

C;Genetics:

A;Gene: ATSP:YUP8H12R.2

A:Map position: 1

[illegible]

RESULT 9  
C96758  
probable protein transporter T18K17.11 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C96758  
R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luro, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C96758  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-539 <STO>  
A:Cross-references: GB:AE005173; NID:g6598860; PIDN:AAF18714.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T18K17.11  
A:Map position: 1

Query Match 16.7%; Score 475; DB 2; Length 539;  
Best Local Similarity 29.8%; Pred. No. 1.2e-27;  
Matches 145; Conservative 82; Mismatches 193; Indels 66; Gaps

QY 87 ATIANFALGIEPCRDVDLQGLQESCLDGEF--SQDYILSTVTEWNLVCEDNWKYP 143  
DB 97 AIVGEASLGLASSE-----WEWIGPKSD---TVVSEWNLLICQHKFLVA 137  
QY 144 LTTSLFFVGVLLGSFVSGQLSDR-FGRKNVLFATMAVQTGFSPLQIFTSWE----- 194  
DB 138 VPSTLFTIGSLFGSGVGYLADSWFGRKKTLTLLSCVL---TFVTAFATFSFNVVYAF 193  
QY 195 -MFTVLVIVGMGOISNYYVAFILGTETLGSVRIIFSTLIGVCTEFAGVMLPLPLFAFY 253  
DB 194 LRFANGFFRSIGS-----CCIVLATEIVGKKRWGVQVGF-FFFTLGLSLPLMAYLE 247  
QY 254 R-DWRMLLLALT-VPGVLCVPLMWFIPESPRWLISORRFEAEDIIQAKAMNNTAVPAV 311  
DB 248 RKSWNLYRIISFPLGVAVCLLPFAFESPRWLLVGRNKEAMVLLKTLARLNGKQLAD 307  
QY 312 IFDSVEELNPLKQKQAFILDFRFRNTAINTMSLLWMLTISVGYFALSIDAENLHGDAY 371  
DB 308 L-SLVDPIDPRDDQTSSEKFKWKVAKRIIIVMMAGFGSGEYVYGIQJNAENLNFNLY 366  
QY 372 LNCFLSALIEIPAYITAWLLRLTPRYITAAVLFWGGVLLFTLQIPVDY----- 423  
DB 367 LTVAVNALMEFPVAFIGSLGLVGNRPLFSNSSILAGFACLLCAVLISHRVIRATISVAK 426  
QY 424 FLISGLVMLGFGTISAFSMLYVFTAELVPLVRNMAVGTSTASRYGSIAPYFVYLGA 483  
DB 427 WLQAVEAGFMASTAYDVLVYVCVELFPPTNVRTAVSLLRQAFMLGASAAPLLVALGR 486  
QY 484 YNRMLPYIVMSGLTVLIGIFITLFPESLIGMTLPETLSOMQKVWFRSGKKTROSMEETEN 543  
DB 487 ESAMMSFIVFGVASVLSGIVSLWLRETRNAPLYETLACQCKA-----BEIEN 533  
QY 544 PKVLIT 549  
DB 534 ETIMIT 539

RESULT 10.  
B86299  
hypothetical protein AAD34691.1 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2000  
C:Accession: B86299  
R:Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; Whit-  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykha-  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti-  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sur-  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: B86299  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-518 <STO>  
A:Cross-references: GB:AE005172; NTD:g4966360; PIDN:AAD34691.1; GSPDB:GN00-  
C:Genetics:  
A:Map position: 1

A:Residues: 1-539 <STO>  
A:Cross-references: GB:AE005173; NID:g6598860; PIDN:AAF18714.1; GSPDB:GNC00141  
C:Genetics:  
A:Gene: T18K17.11  
A:Map position: 1

```
Dd      98 WEWDGSGQKSVISEFGLGCSSLLRGMPSAPYIGAIVGGFFIALIPDDSLGRKKLVLP    15
Qy      175 ATMAVQTGFSELIQSISWENFTVLFIYVGWGQISNYYVAFTILGTETLIGKSVRILESTLG   234
Dd      158 STEFSMI-TSISVFSTNWVIYTFLAKIIFGSRSQTSWSVALVLSERVSRTWR-PRATMI    215
Qy      235 VCTFFAVGYMLPLPFAYFIRD--WMILLALITVPGVLCVPLWWFIPESRWLLSQRRFR     291
Dd      216 PFTLFVLFGMSUGTAFIAQDSMWRYLYLTSPVAVFYCIFLYLALESRWLMHMGDKK        275
Qy      292 EADIIDIKAANKNNATAVDPIFDVSVEELNP LKQQ-----KAFILD,LFTRNAIM         341
Dd      276 EADVLTKMSPKEK-----AYLESWSKLPKQENFEQAPTYSIKDFFEKWAFRRLLVV       330
Qy      342 TIMSLLLWLMTSGSYEALSLDAPNLHGDAVNLCFLSNLETPPAYITAALLRLTLRPXYII      401
Dd      331 MIIMEGL----GISYGYP LAARDIDNVNYLSETLNALVELPTFPVTPILLERNRSSV          386
Qy      402 AAVLFWGG--GWLLPQLPVVDYIFYSIGVMIMGKGFTISA-----FSMLIYVF            447
Dd      387 LVNTLLGGASGYLCFV-----LSTLGTKETAFAFELGTFCARIGFNLMAVE           433
Qy      448 TAELYPTLRNMNAVGVTTSTARSGVISITAPPYFYLGAYNRMPLYIVMGSLTVLIGIEITLFF     507
Dd      434 VMEMPTCYRSSATTMMFQOALVGGACCP LIASIGRIYIPSYSFAIFGTAMSGLMGVLLIL      493
Qy      508 PESLGMTLPETLEQMOK 524
Dd      494 PETKGLSLCDSMEEOEK 510

RESULT 12
B96825
hypothetical protein T8K14_17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2000
C:Accession: B96825
R:Theologian, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; Whit-
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.
ansen, N.P.; Hughes, B.; Huizar, L.
Nature 408; 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khayki-
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maity,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: B96825
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-515 <STO>
A:Cross-references: GB:A0005173; NID:g4835768; PIDN:AND30235.1; GSPPDB:GN00
C:Genetics:
A:Gene: T8K14.17
A:Map position: 1

Query Match                13.4%; Score 382.5; DB 2; Length 515;
Best Local Similarity      26.2%; Pred.No. 8e-21;
Matches 118; Conservative  98; Mismatches 178; Indels 57; Gaps:

Qy . 113 CLD-----GWPESDQDVYSTVTWEHLVCEDNKVKPLTSLTFYGVLL 155
||| ||| :||| |::| ::|::| :|::| ::|::| :|::| ::|::| :|::|
Db 72 CLDTICNPATTDICKIPSRAMDWDGFKSKSVISEDFDECGSSSFSLSPSTFYVGSI 131

Qy 156 GPSVSQLSD-RFRKNVL-FATMAVO-TGEFLOIFSISMETVTLFVIVGMQISNVY 212
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Dd 132 GGVLAMPDGSLGRKQLLFSSFAGMSGISIF-LSSNIW-IYSELKFVIGARSQTQT 189

Qy 213 VAFILGTEILGKSVRIISTLGCVCTFFAVGYMLLPALFYAIR--DWRMLLALTVP-GVL 269
||: |:| :|: |:| :|: |:| :|: |:| :|: |:| :|: |:| :|: |:| :|: |:| :|:
Dd 190 VALVLTISRISTKWR-PRATMVPTFLVIFGFMGSLGIATLVLRHASWKVLYCLSPIAGIH 248
```



C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: P75580  
 R:White, O.; Elsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: P75580  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-454 <WHI>  
 A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PID:AAF12486.1; PID:g646078  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRA0271  
 A:Map position: 2  
 C:Superfamily: yaaU protein

Query Match 11.5%; Score 327; DB 2; Length 454;  
 Best Local Similarity 26.0%; Pred. No. 8.8e-17;  
 Matches 107; Conservative 81; Mismatches 166; Indels 58; Gaps 13;

QY	143	PLTTSLF---	FVGLGSFVSGQLSDRFGKKNVLFATMAYGTGFSFLOIF--SISHEM--	195
DB	67	PAATMLLTATFAGMLFGAWFGYLDVGRRSVELTTVALGVVFGLAGALAPTTLTLLVA	126	
QY	196	-FTVLFIIVGMOQISNYVAFILGILGKSVRIIFSTLGVCTFFFAVGYMLLPFLFAYFI-	253	
DB	127	RELTGFAIGGTLVDYSNMAEFVTANRGRFLVYLES-----FWAGTVVVAALAWWVS	180	
QY	254	-----RDWRMLLALTVPGLVCVPLWFFIPESPRWLSORRFRERAEIDIIQAAKMNNTA	307	
DB	181	TAFAPAEGRWLLGLAALPGVLGLIARIGIPDSRSLARGEAAQARAALQKVAQANGGT	240	
QY	308	VPAVIFDSVEELNPLKQKAPILDLFRTRIAIMTMSLLMLTSGVGFALSILDAPNL-	366	
DB	241	LPNAPL-----AHPDPPRVSPAQLFR-GVLARTPLLMVTFGLSLGYGIFSWLPSEL	294	
QY	367	-----HGDVINCPLSALIEIPAYITAWLLLTLPRLRYIIAAVLFWGG-GVLLFIQLVP	419	
DB	295	RAQGLDLGAVYRSTLLALAQVPGYLLAAYLVEKIGRRVTLVGFLTLGAVGAYLFLAHD	354	
QY	420	VDYVFLSTGLVMLGKFGITSAFSLYVFTAEYPTLVNRNMAVGTSTASRVGSIIPYFV	479	
DB	355	ANTVLLTSALL---\$FALLGAWGSLYAYTPELPTPLRTTGMGLVSGVARLASVWSP---	408	
QY	480	YLGAYNRMLPYIVMGSLTVLIGITFLFFP-----ESLGMTLPETLE	520	
DB	409	SIGA-----MLLTGNTLTALTVPFVCFALAAALAAWGIQVETRQALAEAE	454	

Search completed: January 25, 2002, 10:09:47  
 Job time: 83 sec

Result No.	Query			ID	Description
	Score	Match	Length		
1	2218	78.0	557	OCN2_HUMAN	O76082 homo sapien
2	2165	76.1	557	OCN2_RAT	O70594 rattus norv
3	2164	76.1	557	OCN2_MOUSE	O920e8 mus musculu
4	393	13.8	751	YLX5_CAEEL	P46301 caenorhabdi
5	317.5	11.2	1222	YMP13_CAEEL	Q10947 caenorhabdi
6	307	10.8	435	Y113_CAEEL	Q10917 caenorhabdi
7	294.5	10.4	529	Y0U1_CAEEL	P30638 caenorhabdi
8	290	10.2	400	YCEL_BACSU	O34691 bacillus su
9	274.5	9.6	443	YAAU_ECOLI	P31679 escherichia
10	270	9.5	461	YCSB_BACSU	P46333 bacillus su
11	262	9.2	445	YGC5_ECOLI	O46909 escherichia
12	248	8.7	490	GTRI_CHICK	P46896 gallus gall
13	247	8.7	592	HXT5_YEAST	P38695 saccharomyc
14	245	8.6	459	YDJK_ECOLI	P76230 escherichia
15	245	8.6	566	KHT2_KLUULA	P53387 kluyveromyc
16	244.5	8.6	567	HXT9_YEAST	P40885 saccharomyc
17	243	8.5	451	GTRI_PIG	P20303 sus scrofa
18	243	8.5	492	GTRI_RAT	P1167 rattus norv
19	240	8.4	495	GTRI_CANFA	P47842 canis fami
20	238	8.4	492	GTRI_BOVIN	P27674 bos taurus
21	238	8.4	492	GTRI_MOUSE	P17809 mus musculu
22	237.5	8.3	452	YDJE_ECOLI	P38055 escherichia
23	237	8.3	492	GTRI_HUMAN	P11166 homo sapien
24	237	8.3	546	HXT0_YEAST	P43581 saccharomyc
25	237	8.3	567	YXA4_YEAST	P54862 saccharomyc
26	236	8.3	451	YVAJ_BACSU	P37514 bacillus su
27	234.5	8.2	494	GTRI_SHEEP	P47843 ovis aries
28	234	8.2	491	XYLE_ECOLI	P09098 escherichia
29	231	8.1	522	GTR2_RAT	P12336 rattus norv
30	226	7.9	457	YIRO_YEAST	P40441 saccharomyc
31	225	7.9	492	GTRI_RABIT	P13355 oryctolagus
32	225	7.9	763	RGT2_YEAST	Q12300 saccharomyc
33	223	7.8	413	MUCK_ACICA	P94131 acinetobact

mutation associated with an unconventional splicing abnormality.";  
Biochem. Biophys. Res. Commun. 261:484-487(1999).  
[6]  
RN VARIANT CDSP CYS-211.  
RN MEDLINE=99408248; PubMed=10480371;  
RN R.F.M., Scholte H.R., Rutter J., Hussaarts-Odiik L.M.,  
RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,  
RA Wanders R.J.A.;  
RN "Identification of two novel mutations in OCTN2 of three patients with  
RN systemic carnitine deficiency.";  
RN Hum. Genet. 105:157-161(1999).  
[7]  
RN VARIANT CDSP LEU-478.  
RN MEDLINE=99172075; PubMed=10072434;  
RN Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,  
RA Wang R.J., Fok T.F., Hjeltn M.;  
RN "Mutations of OCTN2, an organic cation/carnitine transporter, lead to  
RN deficient cellular carnitine uptake in primary carnitine deficiency.";  
RN Hum. Mol. Genet. 8:655-660(1999).  
[8]  
RN CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.  
RN MEDLINE=20026865; PubMed=10559218;  
RN Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;  
RN "Mutations in novel organic cation transporter (OCTN2), an organic  
RN cation/carnitine transporter, with differential effects on the  
RN organic cation transport function and the carnitine transport  
RN function.";  
RN J. Biol. Chem. 274:33388-33392(1999).  
[9]  
RN VARIANTS CDSP ARG-283 AND PHE-446.  
RN MEDLINE=20081068; PubMed=10612840;  
RN Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,  
RA Tsuji A.;  
RN "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a  
RN patient with primary systemic carnitine deficiency.";  
RN Hum. Mutat. 15:118-118(2000).  
[10]  
RN VARIANT CDSP LYS-452.  
RN MEDLINE=20145665; PubMed=10679939;  
RN Wang Y., Kelly M.A., Cowan T.M., Longo N.;  
RN "A missense mutation in the OCTN2 gene associated with residual  
RN carnitine transport activity.";  
RN Hum. Mutat. 15:238-245(2000).  
CC -!- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE  
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
CC CARNITINE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,  
CC HEART AND PLACENTA.  
CC -!- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY  
CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE  
CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE  
CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND  
CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL  
CC MYOPATHY OR CARDIOMYOPATHY.  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
CC CATION SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collabora-  
CC tion between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no  
CC modified and this statement is not removed. Usage by and for commer-  
CC cial entities requires a license agreement (See <http://www.isb-sib.ch/announcement>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; AF057164; AAC24828.1; -;  
CC EMBL; AB015050; BAA29023.1; -;  
CC EMBL; AB016625; BAA36712.1; -;  
CC MIM; 603377; -;  
CC MIM; 212140; -;  
CC InterPro; IPR003662; sub.transprtr.  
CC



Db 481 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDITDQMLRVKGMKRRKTPSHR 540  
QY 536 DSMETEENPKVL-ITAF 551  
Db 541 MLKDGQERPTILKSTAF 557

RESULT 2  
OCN2\_RAT OCN2\_RAT STANDARD; PRT; 557 AA.  
AC 070594; OQWLO; 20-AUG-2001 (rel. 40, Created)  
DT 20-AUG-2001 (rel. 40, Last sequence update)  
DT 20-AUG-2001 (rel. 40, Last annotation update)  
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22, MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER) (UST2R) (CT1).  
DE SLC22A5 OR OCTN2.  
GN Rattus norvegicus (Rat).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney.  
RX MEDLINE=98200080; PubMed=9541011;  
RA Schoemig E., Spitzberger F., Engelhardt M., Martel F., Oerding N., Gruendemann D.;  
RT "Molecular cloning and characterization of two novel transport proteins from rat kidney."; FEBS Lett. 425:79-86(1998).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;  
RX MEDLINE=99011422; PubMed=9792817;  
RA Sekine T., Kusuura H., Utsumiya N., Tate N., Tsuda M., Sugiyama Y., Kanai Y., Endou H.;  
RT "Molecular cloning and characterization of high-affinity carnitine transporter from rat intestine."; Biochem. Biophys. Res. Commun. 251:586-591(1998).  
RL [3]  
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=99384224; PubMed=10454528;  
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H., Chen J., Conway S.J., Ganapathy V.;  
RT "Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter."; J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
RL [1-] FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF CARNITINE.  
CC [1-] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC [1-] TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES AND IN THE GLOMERULI IN THE KIDNEY. IN THE MYOCARDIUM, VALVES, AND ARTERIOLES IN THE HEART. IN THE LABYRINTHINE LAYER OF THE PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE BRAIN.  
CC [1-] SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC CATION SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AJ001933; CAA05106.1; -;  
DR EMBL; AB017260; BAA34399.1; -;

DR EMBL: AF110416; AAD54059.1; -;  
DR InterPro: IPR003662; sub.transportr.  
DR Pfam: PF00083; sugar\_tr\_1  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
KW Transport; Transmembrane; Glycoprotein.  
FT TRANSMEM 21 41 POTENTIAL.  
FT TRANSMEM 143 163 POTENTIAL.  
FT TRANSMEM 173 193 POTENTIAL.  
FT TRANSMEM 198 218 POTENTIAL.  
FT TRANSMEM 233 253 POTENTIAL.  
FT TRANSMEM 258 278 POTENTIAL.  
FT TRANSMEM 342 362 POTENTIAL.  
FT TRANSMEM 374 394 POTENTIAL.  
FT TRANSMEM 407 427 POTENTIAL.  
FT TRANSMEM 431 451 POTENTIAL.  
FT TRANSMEM 489 509 POTENTIAL.  
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 275 275 W -> G (IN REF. 2).  
SQ SEQUENCE 557 AA; 62567 MW; 172472E7B0A5F043 CRC64;

Query Match 76.1%; Score 2165; DB 1; Length 557;  
Best Local Similarity 73.2%; Pred. No. 4.9e-134;  
Matches 408; Conservative 68; Mismatches 75; Indels 6; Gaps 3;

QY 1 MRDDEVIAFLGEGPQFORLIFFLLSASIIIPNGFNGMSVVFVLAGTPEHRCRVPDPAANLS 60  
Db 1 MRDDEVIAFLGEGPQFORLIFFLLSASIIIPNGFNGMSVVFVLAGTPEHRCRVPDPAANLS 60  
QY 61 AWRNNSVPLRLRDGREGVPHSCSYRLATIANFSALEPGRDVDLQOLESCLDGWEFS 120  
Db 61 AWRNHSIPLETKDGRQVPSQCRRYRLATIANFSALEPGRDVDLQOLESCLDGWEYN 120  
QY 121 QDVLSTVVTENLVCEENKVPITSLFFVGLGSGVSGQLSDRGRKNVLFATMAVQ 180  
Db 121 KDFVLTSTIVTEMDLVCKDWRAPLTTSLFFVGLMGFSISQGLSDRGRKNVLFATMAVQ 180  
QY 181 TGFSPLOIFSIWEMFTVLFVIVGMGQISNYVAFILGTTELKSGSVRIIFSTLGVCTFFA 240  
Db 181 TGFSPLOIFSVNEMFTVLFVIVGMGQISNYVAFILGTTELKSGSVRIIFATLGVCFYA 240  
QY 241 VGYMLLPFAFIRDRWMLLALVPGVLCVPLWVFPESPRWLISORRPREAEDIIQKA 300  
Db 241 FGFVPLPLFAFIRDRWMLLALVPGVLCVPLWVFPESPRWLISORRPREAEDIIQKA 300  
QY 301 AKMNTAVPAVIED--SVLEELNPLKQKAFILDIFRTRNIAIMTMSLLMLLSVGYFA 358  
Db 301 AKFNCIVAPSTIFDSELDQDLSKPKQSHHIYDLVTRNIRIITIMSLTISVGYFG 360  
QY 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAMLLRLTPRRYIIAIVPWGGVLLFTOLV 418  
Db 361 LSLDTPNLHGDIYVNCFLAEEVPAVYVLAAMLLQHLPRYSISAAALFGGSLVLIQLV 420  
QY 419 PVDYFSLTGLVLMGKFGITSAFMSLYVTFELYPTLVRNMAVGVTSTASRGSIIAPYF 478  
Db 421 PSELYFSLTALYVMGKFGITSAFMSLYVTFELYPTVVRNMGVSVSTASRGSIIAPYF 480  
QY 479 VYLGAYNMLPYVIMGSLTVLIGIFTELPFPEISLGMTLPETLEQMOKVWFRSGK---KTR 535  
Db 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGAPLPDITDQMLRVKGIQWQIQSQTR 540

RESULT 3  
OCN2\_MOUSE OCN2\_MOUSE STANDARD; PRT; 557 AA.  
AC Q920F8; 20-AUG-2001 (rel. 40, Created)  
DT

DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,  
 DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).  
 GN SLC22A5 OR OCTN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=99113835; PubMed=9916797;  
 RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,  
 RA Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsushita T.,  
 RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,  
 RA Tsuji A.;  
 RT "Primary systemic carnitine deficiency is caused by mutations in a  
 RT gene encoding sodium ion-dependent carnitine transporter.";  
 RL Nat. Genet. 21:91-94(1999).  
 RN [2]  
 RN SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.  
 RC STRAIN=C3H;  
 RX MEDLINE=99057546; PubMed=9837751;  
 RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;  
 RA "A missense mutation of mouse OCTN2, a sodium-dependent carnitine  
 RT cotransporter, in the juvenile visceral steatosis mouse.";  
 RL Biochem. Biophys. Res. Commun. 252:590-594(1998).  
 RN [3]  
 RN SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=99384224; PubMed=1045428;  
 RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,  
 RA Chen J., Conway S.J., Ganapathy V.;  
 RT "Functional characteristics and tissue distribution pattern of organic  
 RT cation transporter 2 (OCTN2), an organic cation/carnitine  
 RT transporter.";  
 RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
 CC [-] FUNCTION: SODIUM-ION DEPENDENT, HIGH-AFFINITY CARNITINE  
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
 CC CARNITINE.  
 CC [-] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC [-] DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL  
 CC STEATOSIS (JVS).  
 CC [-] SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
 CC CATION SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 CC EMBL; AB015800; BAA36590.1; -;  
 CC EMBL; AF111425; AAC99787.1; -;  
 CC EMBL; AF110417; AAD54060.1; -;  
 CC MGD; MGI:1329012; SLC22a5.  
 CC InterPro; IPR003662; sub.transprtr.  
 CC Pfam; PF00083; sugar\_trn.1.  
 CC PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
 KW Transport; Transmembrane; Glycoprotein; Disease mutation.  
 FT TRANSMEM 21 41  
 FT TRANSMEM 143 163 POTENTIAL.  
 FT TRANSMEM 173 193 POTENTIAL.  
 FT TRANSMEM 198 218 POTENTIAL.  
 FT TRANSMEM 233 253 POTENTIAL.  
 FT TRANSMEM 258 278 POTENTIAL.  
 FT TRANSMEM 342 362 POTENTIAL.  
 FT TRANSMEM 374 394 POTENTIAL.  
 FT TRANSMEM 407 427 POTENTIAL.  
 FT TRANSMEM 431 451 POTENTIAL.

FT TRANSMEM 489 509 POTENTIAL.  
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 352 352 L -> R (IN JVS).  
 SQ SEQUENCE 557 AA; 62779 MW; 6093F0EB9612B204 CRC64;  
 Query Match 76.1%; Score 2164; DB 1; Length 557;  
 Best Local Similarity 72.9%; Pred. No. 5.7e-134;  
 Matches 406; Conservative 70; Mismatches 75; Indels 6; Gaps 3;  
 QY 1 MDYDEVTAFLGEMGFQRLFFLLSASIIIPNGFNGMSVVFVLAGTPEHRCRVPDAANLSS 60  
 DB 1 MDYDEVTAFLGEMGFQRLFFLLSASIIIPNGFNGMSVVFVLAGTPEHRCRVPDAANLSS 60  
 QY 61 AWRNNSVPLRLRDGREGVPHSCSRYLATIANFSALEPGRDVLQCLQESCLDGWEFS 120  
 DB 61 AWRNNSIPIETKGRQVPQKCRRYRLATIANFSELGLEPGRDVLQCLQESCLDGWEYD 120  
 QY 121 QDVYLTSTVVTENNVCEDNWKVPLTTSILFFVGLGSFVSQSLSDRFGKRVLFATMAVQ 180  
 DB 121 KDVFSTIVTENDLVCKDDWKAPLTTSILFFVGLGMSFISQSLSDRFGKRVLFATMAVQ 180  
 QY 181 TGFSELFQISWEMFTLVFVVGMSQISNVVAFILGTEILGKSVRIIFSTLGVCTFFA 240  
 DB 181 TGFSELFQVFSNFEMFTLVFVVGMSQISNVVAFILGTEILGKSVRIIFATLGVCIFFA 240  
 QY 241 VGYMLLPFAFIRDRMLLALTPVGLVPLVWETPESPRWLISQRRFREAEDIIQKA 300  
 DB 241 FGFVPLPFAFIRDRMLLALTPVGLVPLVWETPESPRWLISQRRFREAEDIIQKA 300  
 QY 301 AKMNNTPAVAVFD--SYEELNPLKQKAFILDIFRTNIAIMTMSLLMLTSGVGYFA 358  
 DB 301 AKINGIVAPSTIFDPSQLNSTKPOLHHYDLIRTNIRVITINSIILWLTISGVYFG 360  
 QY 359 LSIDAPNLHGDAYLNCFLSALIEIPAYITAWLLLTLPRIITAAVLFVWGGVGLFIOLV 418  
 DB 361 LSIDTPNLHGDIYVNCFLAAVEPAVLAALLQYLPRIYSISAAFLGGSVLLFMQLV 420  
 QY 419 PVDYFLSIGLVMLGKFGITSAFSLMYFTAEYPTLVLRNMAVGVTSASRVGSIAPYF 478  
 DB 421 PSELEFLTALVMVGKFGITSAFSLMYFTAEYPTLVLRNMAVGVTSASRVGSIAPYF 480  
 QY 479 VYLGAVNMLPVYVMSLTVLIGITLFFEPESLGMTLPETLEOMQKWKFRSK---KTR 535  
 DB 481 VYLGAYDRFLPYLMGSLTILTALTTLFFPESGVPLPDTIDQMLRVKGIKQWQISQTR 540  
 QY 536 DSMETEENPKVL-ITAF 551  
 DB 541 MQKDGEESPTVLKSTAF 557  
 RESULT 4  
 YLX5\_CAEEL STANDARD; PRT; 751 AA.  
 AC P46501;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE HYPOTHETICAL 84.8 KDA PROTEIN F23F12.5 IN CHROMOSOME III.  
 GN F23F12.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Du 2.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC [-] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

EMBL; U12965; AAA20607.1; -  
DR WormPep; F23F12.5; CE01252.  
DR InterPro; IPR002184; Srb.  
DR InterPro; IPR003662; sub\_transporth.  
DR Pfam; PF02175; Srb; 1.  
DR Pfam; PF00083; sugar\_tr; 1.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 24 44 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT TRANSMEM 102 122 POTENTIAL.  
FT TRANSMEM 137 157 POTENTIAL.  
FT TRANSMEM 200 220 POTENTIAL.  
FT TRANSMEM 235 255 POTENTIAL.  
FT TRANSMEM 320 340 POTENTIAL.  
FT TRANSMEM 348 368 POTENTIAL.  
FT TRANSMEM 379 399 POTENTIAL.  
FT TRANSMEM 410 430 POTENTIAL.  
FT TRANSMEM 432 452 POTENTIAL.  
FT TRANSMEM 515 535 POTENTIAL.  
FT TRANSMEM 547 567 POTENTIAL.  
FT TRANSMEM 583 603 POTENTIAL.  
FT TRANSMEM 614 634 POTENTIAL.  
FT TRANSMEM 678 698 POTENTIAL.  
FT TRANSMEM 751 AA; 84832 MW; A6C4F43540295EFC CRC64;  
SEQUENCE

Query Match 13.8%; Score 393; DB 1; Length 751;  
Best Local Similarity 27.0%; Pred. No. 1.5e-18;  
Matches 119; Conservative 91; Mismatches 184; Indels 46; Gaps 14;

QY 127 TVVTEWNLVCEBN-WKVPVLTSLFFVGVVLGVSFVSQSLDRFGKKNVLPATMAVQGFSP 185  
DB 303 SMVQDFKMFCCQKAYDAAWATIQFVGLGATYHGLDHFGRKPKVPFSGISVGLFGV 362  
QY 186 LQFISISWEMFTLVFVVGMSQSNVYVAFILGTEILGKSVRIIFSLGVCTEFAVGYY-- 243  
DB 363 ASGFAPSWEFVFAFRFVIGHSIASILIVFAYILEFIEPEQRFVLR-----SFFNNGYAR 417  
QY 244 MLLPLFAYFIRDRMMLLALTVPGLCVPLWFPFIPESPRWLISORRFREAEDIIQKAAKM 303  
DB 418 LVFTLACFICGYWRSAAIATSLSLPILPVLLIPESPKWNTKKRPDRARAAEKRVAVL 477  
QY 304 NNTAVPAV-----IFDSVEELNPLKQOKAFIL-DLFRTRNIAITMISLLMLTSV 354  
DB 478 --SGIPYNDQDQIEISEKLEE---KSTKIYTMKDLFTSWIATYRTIVVGSLSFSTSL 531  
QY 355 GYFALSIDAPNLGHDAVNCFLSALIEPAYITAWLLRLTP---RRIYI-----IAAVLF 406  
DB 532 SAFGSDNLNGNAGNFVLSQFVSAVTAFAKIFVFLDLTVVPSDRRLHQYQIAMILC 591  
QY 407 WGGVGLLFQIOLPV-----DYFLSLGLVLMGKFGITSAFSMLYVFTAEIYPLVR 457  
DB 592 Y--CVIMVLMILPESDCSGSQRDLAIITINIGVSFEIT--WDACYLVAVECFPKIR 647  
QY 458 NNAVGVTSASRGVSIATPYVYVYGANRMLPYIYVMSL-TVLGIGITLFPESLGMTLP 516  
DB 648 TIGITCSLLARTGALLAPMAYISDIYRPAPYAVWCISGTSISLISCVLPLDPTKGVDL- 706  
QY 517 ETLEQMKVWFRSGKTRD 536  
DB 707 AALDPTTELDYRKKSMTEN 726

RESULT 5

YMP3\_CAEEL  
ID YMP3\_CAEEL STANDARD; PRT; 1222 AA.  
AC Q10947; 1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 139.9 KDA PROTEIN B0361.3 IN CHROMOSOME III.  
GN B0361.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Du Z.;  
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases  
RL -J- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

EMBL; U00031; AAK18964.1; -  
DR WormPep; B0361.3; CE00752.  
DR Pfam; PF00083; sugar\_tr; 1.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 51 71 POTENTIAL.  
FT TRANSMEM 148 168 POTENTIAL.  
FT TRANSMEM 182 202 POTENTIAL.  
FT TRANSMEM 232 252 POTENTIAL.  
FT TRANSMEM 263 283 POTENTIAL.  
FT TRANSMEM 307 327 POTENTIAL.  
FT TRANSMEM 329 349 POTENTIAL.  
FT TRANSMEM 355 375 POTENTIAL.  
FT TRANSMEM 382 402 POTENTIAL.  
FT TRANSMEM 413 433 POTENTIAL.  
FT TRANSMEM 442 462 POTENTIAL.  
FT TRANSMEM 1222 AA; 139868 MW; CBA42A80A254FB80 CRC64;  
SEQUENCE

Query Match 11.2%; Score 317.5; DB 1; Length 1222;  
Best Local Similarity 20.3%; Pred. No. 2e-13;  
Matches 114; Conservative 106; Mismatches 219; Indels 123; Gaps 14;

QY 3 DYDEVIAFTLGEWGP-----QRLIFPLLSASIPNGFNGMSVYFLAGTPEHRCRVPD 54  
DB 35 DPKDFEAYGAYGKYQIFTYVLVQTLNPFYSSSMVI-----MSFVL--NLEKQCE--- 83  
QY 55 AANLSAWNNNSVPLRLRGVPHSCSYRLATIANFALGEPGRVDVLGQLEQESCL 114  
DB 84 -----YKNETIP-----ISETCQ-----IETESKAFNLNGEYC- 113  
QY 115 DGWEFSODVYL-----STVVTENNLVCEDNWKVPLTTSFFVGVLLGSEFVSQSLDR 166  
DB 114 ---GIANTLVNVTNOKASTNLLVDLSCSHWFQEFGLTFTIGAVIATVFMMLADR 170  
QY 167 FGRKNYLFATMAVQGFSLQIFSIISWEMFTVLFVVGMSQISNYYVAFILGTEILGKSV 226  
DB 171 YGRKPIIVTTAILAFLANNAAGSPNFAIFLILRAFIGACSDSYLSVASVATCEYLSEKA 230  
QY 227 RIIFSLGVCTEFAVGYMLLPFAFYIRDRMMLLALTVPGLCVPLWFPFIPESPRWLIS 286  
DB 231 R-AWITVYVYVAVSLGMMVWTLVLTMTDWMRYFYIVSLPGVYGFMALTFPESPHWLIT 289  
QY 287 ORFREAEADIIOKAAKMNNTAVPAVIFDSVEELNPLKQOKAFILDLFRTRNIAITMISL 346  
DB 290 KNKTEKLKVIKTANRM-----VISL 310

QY 347 LLMLTSGVGFALSDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTLPVRYIIAAVL 406  
 Db 311 V-----YFAISFMSVGLGDOVQAFLYSSLIEIPAGLAVIPLMKWGRKMIVTWCLV 362  
 QY 407 WGG-----GVLLFTQVDPVYFISGLVLMGKFGITSAFSMLYVFTAEIYPTLVRNMAVG 462  
 Db 363 FQTLALIGVTVFL-----DSYEFKVLMLVAKVMATIIYSVHPWATEQFPTSVRSCLFS 417  
 QY 463 VTSTASVSGSIAPFYVYLGAYNRMPLPYIVMGSLTVLIGIFTFFPESLGMILPETLEQM 522  
 Db 418 LMIIPQSGMIIMPYKVIHIVSNWIPFVVVIALFSPISATLAPMLHETKKNKLPDIEDSL 477  
 QY 523 -----QKVKFRSGKKTDSM 538  
 Db 478 SYPSETNDLSAYRRSKSSSSV 499

# RESULT 6 YTI3\_CAEEL

ID YTI3\_CAEEL STANDARD; PRT; 435 AA.  
 AC Q10917;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE HYPOTHETICAL 48.6 KDA PROTEIN B0252.3 IN CHROMOSOME II.  
 GN B0252.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Du 2., Waterston R.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U23453; AAC46757.1;  
 DR WormPep; B0252.3; CE02419.  
 DR InterPro; IPR003662; sub\_trnsportr.  
 DR Pfam; PF00083; sugar\_tr; 1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 55 75 POTENTIAL.  
 FT TRANSMEM 384 404 POTENTIAL.  
 SQ SEQUENCE 435 AA; 48571 MW; 5F6160359FA1840B CRC64;

Query Match 10.8%; Score 307; DB 1; Length 435;  
 Best Local Similarity 23.5%; Pred. No. 3.3e-13;  
 Matches 94; Conservative 73; Mismatches 155; Indels 78; Gaps 8;  
 QY 128 VYTENLVCECDNWKVPLTSLFFVGLGVSEVSGQLSDRGKKNVLFATMAVQGFSEFLQ 187  
 Db 91 VADEFDLTGDSALAEISTTFVWVGNMIGMFIPPLADHYGLRFLPVFVATVLLMVGGMIS 150  
 QY 188 IPSISWEMFTVLVLVGMGQISNYVVAEILTEILGKSVRIIFSTLGVCTFFAVGYMLLP 247  
 Db 151 AFSISIMMFCIMRMTHGIFYTAAGLAGVGLGYENTPLRLR-FFTSVYFGVMVWVYACFLG 209  
 QY 248 LFAYFIQWRLMLLALTPGV-LCVPLWVFPESPRWLISORREREADIQK-AAKWN 305  
 Db 210 LLAYILPWRVLMCSIPNIFVALLIYMTVPESLHFLVSSQNEKTEAWLEKIRGPKGD 269  
 QY 306 TAVPAVIFSDSVEELNPLK-----QOKAFILDIFRTRNIAIMTMSLLWMLTSSVGYFALS 361

Db 270 ISASDIVEDRNGSSFKTLCREIKFTFTLFO----- 302  
 QY 362 DAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTLPVRYIIAAVLFWGGVLLFIQLVPVD 421  
 Db 303 -----DRRYI-----YLFWIGILLYEGRKPL- 324  
 QY 422 YFELSIGLVLMGKFGITSAFSMLYVFTAEIYPTLVRNMAVGVTASVSGSIAPFYVYL 481  
 Db 325 FEFC-----HEFGRSSSLHFFSDFHEQIEFTDGRNKCIGFCETLSRFGMLSPYLSHL 378  
 QY 482 GAYNRMPLPYIVMGSLTVLIGIFTFFPESLGMILPETLEQ 521  
 Db 379 TAVHALAPAITLSLAVSGGLTLILPETLNTKLPSTIAE 418

# RESULT 7 YOIU\_CAEEL

ID YOIU\_CAEEL STANDARD; PRT; 529 AA.  
 AC P30638; O21101;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III.  
 GN ZK637.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,  
 RA Halloran N., Green P., Thierly-Mieg J., Qiu L., Dear S., Coulson A.,  
 RA Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,  
 RA Ainscough R., Waterston R.;  
 RL "The C. elegans genome sequencing project: a beginning."  
 RL Nature 356:37-41(1992).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=BRISTOL N2;  
 RA Durbin R.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; Z11115; CAA77460.1;  
 DR EMBL; Z22175; CAA77460.1; JOINED.  
 DR EMBL; Z22175; CAA80131.1;  
 DR EMBL; Z11115; CAA80131.1; JOINED.  
 DR PIR; S15786; S15786  
 DR WormPep; ZK637.1; CE06638.  
 DR InterPro; IPR003662; sub\_trnsportr.  
 DR Pfam; PF00083; sugar\_tr; 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 KW Hypothetical protein; Transmembrane; Transport.  
 FT TRANSMEM 86 106 POTENTIAL.  
 FT TRANSMEM 122 142 POTENTIAL.  
 FT TRANSMEM 158 178 POTENTIAL.  
 FT TRANSMEM 238 258 POTENTIAL.  
 FT TRANSMEM 320 340 POTENTIAL.  
 FT TRANSMEM 373 393 POTENTIAL.  
 FT TRANSMEM 411 431 POTENTIAL.  
 FT TRANSMEM 482 502 POTENTIAL.









Db	393	VLAQGMQVILLILLATVLLVGFVVTLWAPETKALPL	429
RESULT	12		
GTSL_CHICK			
ID	GTSL_CHICK	STANDARD;	PRT; 490 AA.
AC	P46896;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	03-NOV-1997 (Rel. 35, Last annotation update)		
DE	GLUCOSE TRANSPORTER TYPE 1 (GT1).		
GN	SUC2AI OR GLUT1.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96157892; PubMed=8589457;		
RA	Wagstaff P., Kang H.Y., Mylott D., Robbins P.J., White M.K.;		
RT	"Characterization of the avian GLUT1 glucose transporter:		
RT	differential regulation of GLUT1 and GLUT3 in chicken embryo		
RT	fibroblasts."		
RL	Mol. Biol. Cell 6:1575-1589 (1995).		
CC	- - FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. THIS ISOFORM MAY BE		
CC	RESPONSIBLE FOR CONSTITUTIVE OR BASAL GLUCOSE UPTAKE. HAS A VERY		
CC	BROAD SUBSTRATE SPECIFICITY; CAN TRANSPORT A WIDE RANGE OF ALDOSES		
CC	INCLUDING BOTH PENTOSE AND HEXOSE (BY SIMILARITY).		
CC	- - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.		
CC	- - SIMILARITY: HIGH TO OTHER GLUCOSE TRANSPORTERS.		
CC	- - SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC			
DR	EMBL; L07300; AAB02037.1;		
DR	InterPro; IPR003663; Sugar transportr.		
DR	InterPro; IPR003662; sub_transportr.		
DR	Pfam; PF00083; sugar_tr; 1.		
DR	PRINTS; PRO0171; SUGTRNSPORT.		
DR	PRINTS; PRO0172; GLUCTRNSPORT.		
DR	PRINTS; PRO1190; GLUCTRSPORT1.		
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.		
DR	PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.		
KW	Transmembrane; Sugar transport; Transport; Glycoprotein;		
KW	Multigene family.		
FT	TRANSMEM	12	POTENTIAL.
FT	DOMAIN	33	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	33	POTENTIAL.
FT	TRANSMEM	66	POTENTIAL.
FT	TRANSMEM	95	POTENTIAL.
FT	TRANSMEM	126	POTENTIAL.
FT	TRANSMEM	146	POTENTIAL.
FT	TRANSMEM	155	POTENTIAL.
FT	TRANSMEM	175	POTENTIAL.
FT	TRANSMEM	185	POTENTIAL.
FT	DOMAIN	205	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	206	POTENTIAL.
FT	TRANSMEM	271	POTENTIAL.
FT	TRANSMEM	326	POTENTIAL.
FT	TRANSMEM	337	POTENTIAL.
FT	TRANSMEM	357	POTENTIAL.
FT	TRANSMEM	368	POTENTIAL.
FT	TRANSMEM	401	POTENTIAL.
FT	TRANSMEM	429	POTENTIAL.
FT	TRANSMEM	449	POTENTIAL.
FT	DOMAIN	450	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	44	N-LINKED (GLCNAC... ) (POTENTIAL).
SQ	SEQUENCE	490 AA;	54086 MW; 63BB176812D06104 CRC64;
Query Match		8.7%;	Score 248; DB 1; Length 490;

Best Local Similarity	24.2%;	Pred. No. 2.6e-09;	
Matches 117; Conservative	87;	Mismatches 170;	Indels 110; Gaps 21;
QY	117	WFESQDVYLSVTWTEWNLVCEDNWKPVLTTSLFFVGVLLGSFVSQOLSDRGRK	----- 170
Db	49	YRYEERIPATLTLWS	-----LSVAIFSVGGMIGSFVGLFVNRGRRNMLMS 98
QY	171	NVLFATMAVOTGFSFLQIFSISWEM	-----FTVLFVIVGMGOISNYVAFIL 217
Db	99	NILAFALAAVLMGFSKM	-----ALSFEMILILGRFTIIGLYSLTTGTFVPMYGEVSPALRGAL 155
QY	218	GTEILGKSVRIIFSTLGVCTFAVCYMLLPFAFI	---RD--WRMLLLALTVPGLVCPV 272
Db	156	GP	-----FHOLGI-----VLGILTAQVFGDLNMGDSLMLPLLLGFIIVFALLQCI 201
QY	273	LWVFIPESPRW-LISORFREAEADIIQAKARNNTAVPAVIFDSVEELNP-LKOOKAFIL	330
Db	202	ILFPAPESPPELLINNEENAKSVLKKL	--RGITDVSDDLQEMKEESROMKREKKVTIM 259
QY	331	DLFRTRN	-----IAIMTMSLLMLTSVGFALSADPNLHGDAYLNCFLSALIEIPA 384
Db	260	ELFRSPMYRQPIILIAIVLQLSQSGINAVFYSTSI	-----FEKSGVEQPV 306
QY	385	YIT	-----AWLLRLTLPRR--YIIAAVLFWGGVLLFIQIOLVPDYY---FL 425
Db	307	YATIGSGVVNTAFTVVSILFVVERAGRRTLHLIGLAGMAGCAILMTIATLTDQPMWSYL	366
QY	426	SIGLVMKFGITSAFS	-----MLYVFTAELPTLVNRMAVGTSTASRVGS-IIAPYFV 479
Db	367	SIVAI	-----FGVAFGEIGPGPIPFWIVAEELFSQGPRAAFVAGLSNWTSNFIVGMGFQ 422
QY	480	YLGAYNRMPLPYVMGSLTVLIGITLIF-FPESLGMTLPETLEQMKVWFPSGKKTRDSM	538
Db	423	YIAQLGSGYVFIETVLVLVLFIFTYFKVPETKGTDEIAYR	-----PROGASQSDK 476
QY	539	ETEE 542	
Db	477	TPDE 480	
RESULT	13		
HXT5_YEAST			
ID	HXT5_YEAST	STANDARD;	PRT; 592 AA.
AC	P38695;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	01-FEB-1995 (Rel. 31, Last annotation update)		
DE	PROBABLE GLUCOSE TRANSPORTER HXT5.		
GN	HXT5 OR YHR096C.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MC996;		
RA	Reifenberger E., Koetter P., Ciriacy M.;		
RL	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S288C / AB972;		
RX	MEDLINE=94378003; PubMed=8091229;		
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,		
RA	Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,		
RA	Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,		
RA	Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,		
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,		
RA	Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,		
RA	Vaudin M.;		
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome		
RT	VIII."		
RL	Science 265:2077-2082(1994).		
CC	- - FUNCTION: PROBABLE GLUCOSE TRANSPORTER.		



CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- MISCELLANEOUS: GLUCOSE TRANSPORT IS THOUGHT TO BE MEDIATED BY TWO  
 CC KINETICALLY DISTINCT SYSTEMS, A GLUCOSE-REPRESSIBLE HIGH-AFFINITY  
 CC SYSTEM AND A CONSTITUTIVE LOW-AFFINITY SYSTEM.  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X77961; CAAS4923.1; -.  
 CC EMBL; U00060; AAB68934.1; -.  
 CC PIR; S43742; S43742.  
 CC PIR; S46726; S46726.  
 CC SGD; S0001138; HXT5.  
 CC InterPro; IPR003663; Sugar\_transprtr.  
 CC InterPro; IPR003662; sub\_transprtr.  
 CC Pfam; PF00083; sugar\_tr; 1.  
 CC PRINTS; PR00171; SUGTRNSPORT.  
 CC PRINTS; PR00172; GLUCTRNSPORT.  
 CC PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
 CC PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
 CC Repeat; Transmembrane; Sugar transport; Glycoprotein.  
 KW DOMAIN 1 82 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 83 103 POTENTIAL.  
 FT DOMAIN 104 137 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 138 158 POTENTIAL.  
 FT DOMAIN 159 164 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 165 185 POTENTIAL.  
 FT DOMAIN 186 195 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 196 216 POTENTIAL.  
 FT DOMAIN 217 222 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 223 243 POTENTIAL.  
 FT DOMAIN 244 257 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 258 278 POTENTIAL.  
 FT DOMAIN 279 380 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 381 401 POTENTIAL.  
 FT DOMAIN 402 409 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 410 430 POTENTIAL.  
 FT DOMAIN 431 450 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 451 471 POTENTIAL.  
 FT DOMAIN 472 487 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 488 508 POTENTIAL.  
 FT DOMAIN 509 512 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 513 533 POTENTIAL.  
 FT DOMAIN 534 592 EXTRACELLULAR (POTENTIAL).  
 FT SEQUENCE 592 AA; 66251 MW; 6906721BAC1A5F87 CRC64;  
 Query Match 8.7%; Score 247; DB 1; Length 592;  
 Best Local Similarity 22.9%; Pred. No. 3.6e-09;  
 Matches 114; Conservative 70; Mismatches 172; Indels 142; Gaps 22;  
 Qy 124 YLSTVTVENLVCDNKKVPLTSLFFVGVLLGSFVSQGLSDRGKKNVLFATMAVQTGF 163  
 Db 130 YLSDV-----RTGLMVSIFNGCAIGGIVLKLGMGRKIGLMTVVVYISG 177  
 Qy 184 SFLQIFST-SWEMFTLVFVVGMOISNYYVAFILGTEILGKSVRIIFSTLGVCTFFAVG 242  
 Db 178 IIIQIASIDKKWQYFIRIGIISGLGVGIGITVLAPMLISEVSPKQLR---GLVSC----- 228  
 Qy 243 YMLPLFAFYFIR-----DWRMLLLALTVPGLVCLPLMW-----FIPE 279  
 Db 229 YQLMTTFGIFLGYCTNFGTKNSNSVQWR-----VPLGLCFA-WSIFMIVGMFTVPE 279  
 Qy 280 SPRLMISORRPREEDIIQAKKNNTAVPAVFD-----SVE-----ELNPL 322  
 Db 280 SPRLVLEVKGTEAKRSLRANKTTDS-PLVTLEMYNQSSIEAERLAGSASWGLVTVG 338

Qy 323 KQKAFILDFPRTRNIAITMTSLMLLWMLTSVGYF-----ALSDAPNLHGDAYLN 373  
 Db 339 KPQ-----MFR--TLGMWMIQSLOQLGDNFFYYGTTIFQAVGLE-----DSFET 383  
 Qy 374 CFLSALIEIPAYITAWLLRLPRRYIAAFLFWG--GGVLLFIQLVPVDYFSLISGLVM 431  
 Db 384 AIVLGVVN--FVSTFFSLYIVD-RFGRNCLLWGCVMICCYV-----VVASGVTR 432  
 Qy 432 LKRGF-----ITSAFSMLYVFTAEIPLTVLRNMAVGVTTSTA 467  
 Db 433 LWPNGQDQSPSSKGAGNCMIVFACFYIFCFATWAPVAVYLISESYPLVRGKAMSIASAC 492  
 Qy 468 SRVGSIIAPYFV--YLGAYNRMLPYIVMGSLLTVLIGITFTLFFPESLGMTLPETLEQMOK- 524  
 Db 493 NWINGFLISLFTPTTISAINFYGYGVFNCMVFAFYVFFVFPETKGLTLEVNMEYEN 552  
 Qy 525 -----VKWFRSGKKTRD 536  
 Db 553 VLPWKSTKWIPPSRRTD 570  
 RESULT 14  
 YDJK\_ECOLI  
 ID YDJK\_ECOLI STANDARD; PRT; 459 AA.  
 AC P76230; P76911;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YDJK.  
 GN YDJK OR B1775.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RA "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Samedai G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;  
 RA "A 570-kb DNA sequence of the Escherichia coli K-12 Genome  
 RT Corresponding to the 28.0-40.1 min Region on the Linkage Map.";  
 RL DNA Res. 3:363-377(1996).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE000272; AAC74845.1; -.  
 CC EMBL; D90821; BAA15573.1; ALT\_INIT.  
 CC EcoGene; EGI3487; ydjk.  
 CC InterPro; IPR003662; sub\_transprtr.

DR Pfam; PF00083; sugar\_tr; 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; FALSE NEG.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; FALSE NEG.  
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane;  
 KW Complete proteome.  
 FT TRANSMEM 26 46 POTENTIAL.  
 FT TRANSMEM 61 81 POTENTIAL.  
 FT TRANSMEM 91 111 POTENTIAL.  
 FT TRANSMEM 113 133 POTENTIAL.  
 FT TRANSMEM 154 174 POTENTIAL.  
 FT TRANSMEM 182 202 POTENTIAL.  
 FT TRANSMEM 272 292 POTENTIAL.  
 FT TRANSMEM 302 322 POTENTIAL.  
 FT TRANSMEM 330 350 POTENTIAL.  
 FT TRANSMEM 352 372 POTENTIAL.  
 FT TRANSMEM 400 420 POTENTIAL.  
 FT TRANSMEM 421 441 POTENTIAL.  
 SQ SEQUENCE 459 AA; 49602 MW; BC9AB53E8BDD77 CRC64;

Query Match 8.6%; Score 245; DB 1; Length 459;  
 Best Local Similarity 23.4%; Pred. No. 3.7e-09;  
 Matches 97; Conservative 81; Mismatches 177; Indels 60; Gaps 17;

QY 134 LVCDN-----WKVPLTSLF-----FVGLLGSFVSQSDREFGRKNVLF 175  
 DB 34 LVCNNAVAGGLILAQKALGWTDNSTATFSALTATAGMFGALVGLIGDKTGRNAFTL 93  
 QY 176 TMAVQTFGSLQIFSTSWEMFVLFVIVGMQISNVVAFILGTEIL-GK-----SVRII 229  
 DB 94 YEAIHTASVMVGFSPNMFDFLIACRFVGVGLGALLVTLFAGFTFYMPCGRNRTSSRSVS 153  
 QY 230 FS---TLGVCTEFVAVGMLLPLFAFIRDRWMLLALTVPGLVCPPL-WWFIPESPRLII 285  
 DB 154 FIGNWSPYPLCSLIANG--LTPLIS-AENWRVQLLIPALISLIATLALAWRPESPRLIE 210  
 QY 286 SQRRPREAEDITQK-----AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTRNI--- 338  
 DB 211 SGRYQAEAKVMSIEBEGVIRGTGKPLPPVVIADDBKPAQVAPYSALLTGVLLKRVILGS 270  
 QY 339 AIMTIMSLLLMLTSGVYFALSIDAPNLHGDAYLNCFLSALIEIP-AYTAWLLLRTPR 397  
 DB 271 CVLIANNVQVYTLINWLPFIEMTQGLNKDSIVLNTM--SMFGAPFGFIAMLVMDKIPR 328  
 QY 398 RYIIAAVFWGGVLLFIQLVPVDYFYS--TGLVNLCKFGTISAFSMLY-----VFTA 449  
 DB 329 KTM-----GVGLILIAVLGYIYSLOTSMLLITLIGFFLIT--FVYMVVCYASAVVP 379  
 QY 450 ELYPTLVNRNAVGVSTASRVGSIIAPY--FVYLGAYNRMPLPYIVGSLTVLIGI 502  
 DB 380 EIWPTKLRGLSGLANAVGRISGIAAPYANAVALLSSYGVTVGTFILLGAVSIVAI 434

RESULT 15  
 KHT2\_KLULA  
 ID KHT2\_KLULA STANDARD; PRT; 566 AA.  
 AC P53387;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HEXOSE TRANSPORTER 2.  
 GN KHT2.  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAG6;  
 RX MEDLINE=98028406; PubMed=9363776;  
 RA Weirich J., Goffinl P., Kuger P., Ferrero I., Breunig K.D.;  
 RT "Influence of mutations in hexose-transporter genes on glucose  
 repression in Kluyveromyces lactis."

RL Eur. J. Biochem. 249:248-257(1997).  
 CC !- FUNCTION: PROBABLE GLUCOSE TRANSPORTER.  
 CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC !- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z47080; CAA87389.1; -;  
 DR InterPro; IPR003663; sugar\_transprot.  
 DR Pfam; PF00083; sub\_transprot.  
 DR PRINTS; PR00171; SUGTRNSPORT.  
 DR PRINTS; PR00172; GLUCTRNSPORT.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
 KW Repeat; Transmembrane; Sugar transport; Transport; Glycoprotein.  
 FT TRANSMEM 58 78 POTENTIAL.  
 FT TRANSMEM 114 134 POTENTIAL.  
 FT TRANSMEM 141 161 POTENTIAL.  
 FT TRANSMEM 169 189 POTENTIAL.  
 FT TRANSMEM 199 219 POTENTIAL.  
 FT TRANSMEM 234 254 POTENTIAL.  
 FT TRANSMEM 357 377 POTENTIAL.  
 FT TRANSMEM 386 406 POTENTIAL.  
 FT TRANSMEM 429 449 POTENTIAL.  
 FT TRANSMEM 465 485 POTENTIAL.  
 FT TRANSMEM 491 511 POTENTIAL.  
 SQ SEQUENCE 566 AA; 62727 MW; 9784173EC2375444 CRC64;

Query Match 8.6%; Score 245; DB 1; Length 566;  
 Best Local Similarity 22.3%; Pred. No. 4.7e-09;  
 Matches 106; Conservative 78; Mismatches 196; Indels 98; Gaps 17;

QY 139 NWKVPILTSLFFVGVVLGSFVSQSDREFGRKNVLPATMAVQTFGSLQIFSI-SWEMFT 197  
 DB 109 NVRTGLVIFNIGCAIGGIILSKLGDYGRRIGLMTWLVVYVVGIIQIASIDKWKVQYF 168  
 QY 198 VLFVIVGMQISNVVAFILGTEILGKSVRIFTSLGVCTFFAVGVMLLPLEAVFIR--- 254  
 DB 169 IGRISGLGVGGISVLSFPLISETAPKHIR-----GLVSP---YQLMITFGIFLYCYT 219  
 QY 255 -----DWRMLLALTVPGLVCPVLPWFIPESPRWLISQRRPREAEDIIQKAAK 303  
 DB 220 NYGKTYSNSVQWRVPLGLCFAWAIFMTGMLFVPESPFLVEKDRIDEAKRSIAKSNKV 279  
 QY 304 N-NTAVPA--VIFDSVEELNPLKQKAFILDLFRTR-NIAIMTMSLLL---WMLTSV 354  
 DB 280 SYEDPAVQAEVDLICAGVEA--ERLAGSASIKELFSTKTKVFORLIMGMLIQSFOOLTGN 337  
 QY 355 GYF-----ALSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRTPRPIIAAVL 405  
 DB 338 NYFPYVGTTFISVSGMD-----DSFETSVLGVINFASTFVAIYVVDKGRR---KCL 387  
 QY 406 FWGGVLLFTQLVPVDYFYSIGLVMLGKFG-----ITS 439  
 DB 388 LWGAAMTACMVV-----FASVGVTRLPDGDANHPETASKAGNCNMFACFVIFCFATS 442  
 QY 440 AFSMLYVFTAEYPTLVNRNAVGVSTASRVGSIIAPYFV--YLGAYNRMPLPYIVGSLT 497  
 DB 443 WAPIAYVVAESYPLRVKAKMAIATASWINGFLNGFTPTTSAIHFYGYGVFMGCLV 502  
 QY 498 VLIGITFLFPFESLGMTLPTEQMOK-----VKWFRSGKTR---DSMETEENP 544  
 DB 503 AMFFYVFFVFPETKGLTLEEVQEMWEEGVLPWKSSSWPSSRRNAGYDVALQHDKEP 560

Search completed: January 25, 2002, 10:15:19  
Job time: 325 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 10:09:19 ; Search time 47.78 Seconds  
(without alignments)  
1686.816 Million cell upd

Title: US-09-521-195-1  
 Perfect score: 2845  
 Sequence: 1 MRDYDEVIAFLGWGPPFRL.....KKTRDSMTETENPKVLITAF 551

Scoring table: BLOSUM62

scoring table: `ELCSCMOZ`  
 Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length:	200000000
Maximum DB seq length:	200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

RESULT 1

3: sp\_lung:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	2845	100.0	551	4	O14546		O14546 homo sapien
2	2833	99.6	551	4	O9H015		O9H015 homo sapien
3	2484	87.3	553	11	Q9R141		Q9R141 rattus norv
4	2470	86.8	553	11	Q9Z306		Q9Z306 mus musculus
5	2034.5	71.5	564	11	Q9WRN6		Q9WRN6 mus musculus
6	857	30.1	548	5	O9VCA2		O9VCA2 drosophila
7	852	29.9	548	5	O01384		O01384 drosophila
8	819	28.8	567	5	O9VCA3		O9VCA3 drosophila
9	773.5	27.2	561	5	O9V616		O9V616 drosophila
10	754.5	26.5	568	5	Q9U539		Q9U539 caenorhabdi
11	754.5	26.5	576	5	O02270		O02270 caenorhabdi
12	707	24.9	554	6	O77504		O77504 cryotolagus
13	703	24.7	555	4	O15244		O15244 homo sapien
14	700	24.6	554	6	O02713		O02713 sus scrofa
15	699.5	24.6	593	11	P70485		P70485 rattus norv
16	696.5	24.5	553	11	O70577		O70577 mus musculus
17	695.5	24.4	555	11	Q9R0W2		Q9R0W2 rattus norv
18	695	24.4	674	5	O9V1K2		O9V1K2 drosophila
19	692	24.3	556	11	O08966		O08966 mus musculus

Query Match  
100.0%; Score 2845; DB 4; Length 551;

Best Local Similarity 100.0%; Pred. NO. 8.3e-167;  
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRDYDEVIAFLGEWGPFQRLIFFELLSASIIPNGFNGMSVFLAGTPEHRCRVPDAANLSS 600  
QV

1 MRDYDEVIAFLGEWGPFORLIFFLLSASIIIPNGFNGMSVFLAGTPEHRCRVDPDAANLSS

OV 61 AWRNNSVPLRLRDGREVPHPHSCSRYRLATIANFSA LGLEPGRDVPDLGOLEOESC LDGWEEFS 120

61 AWRNNSVPIRLRDGREVPHSCSRVRIATIANFSAIGLEPGRDVGLEI.EQESCLDGWEFS 120



```
QY 121 QDYVLSVTWTEWNLVCEENKWKPLTTSLSFFVGVLLGSFVSGQLSDREGRKNVLFATMAVQ 180
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 KDYFLSTIVTEWNLVCEENKWKPLTTSLSFFVGVLLGSFVSGQLSDREGRKNVLFATMAVQ 180
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 TGFSLQIFISWEMFTVLFVIVGMQISNYVAVFTLGTTEILGKSVRIIFSTLGVCTFFA 240
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 TGFSLQIFISWEMFTVLFVIVGMQISNYVAVFTLGTTEILGKSVRIIFSTLGVCTFFA 240
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 VGYMLPLFAFYFIRDRWMLLALATVPGVLCVPLWMTIPESPRWLISQRRFREAEIIOKA 300
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 IGYMVLPLFAFYFIRDRWMLLALATVPGVLCVPLWMTIPESPRWLISQRRFREAEIIOKA 300
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 AKMNTAVPAVIEDSVF--ELNPLKQOKAFILDLFTRNIAIMTMSLLWMLTSVGYFA 358
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 301 AKMNTAVPAVIEDSVF--ELNPLKQOKAFILDLFTRNIAIMTMSLLWMLTSVGYFA 358
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAAVLFWGGVLLFIQIV 418
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 361 LSLNPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAGVLFWGGVLLFIQIV 420
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 419 PVDYFSLGVLMLGKFGITSAFMSLYVFTAEYPTLVNMAVGVTSASRVGSIITAPYF 478
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 421 PEDYFVSLGVLMLGKFGITSAFMSLYVFTAEYPTLVNMAVGVTSASRVGSIITAPYF 480
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 479 VYLGAYNRMPLPYIVMGSLTVLIGLFTLFPESLGMTLPETLEOMQVKWFRSGKKTDRSM 538
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 481 VYLGAYNRMPLPYIVMGSLTVLIGLFTLFPESLGMTLPETLEOMQVKWFRSGKKTDRSM 538
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 539 ETEENPKVLITAF 551
:||||:||||:
Db 541 DREENPKVLITAF 553
:||||:||||:
RESULT 4
Q92306 PRELIMINARY; PRT; 553 AA.
ID Q92306 AC Q92306;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE ORGANIC CATION TRANSPORTER.
GN SLC22A4 OR OCTN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL 6J; TISSUE=KIDNEY;
RA Nezu J.;
RT "Mouse OCTN1 : Polyspecific organic cation transporter.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AB016257; BAA36626.1;
DR MGI; MGI:1353479; SLC22a4.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 553 AA; 62290 MW; C4D66BC061398653 CRC64;
```

```
Query Match 86.8%; Score 2470; DB 11; Length 553;
Best Local Similarity 84.6%; Pred. No. 6.4e-144;
Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

QY 1 MRDYDEVIAFLGEGWGPFORLIFFLLSASIIIPNGFNGMSVVFVLAGTPHRCRVPDPAANLSS 60
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MRDYDEVIAFLGEGWGPFORLIFFLLSASIIIPNGFNGMSVVFVLAGTPHRCRVPDPAANLSS 60
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 AWRNNVPLRLRDGREVPHSCSRVRLATIANFSALGLEPGRDVLGQLESCLDGWEYS 120
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
Db 61 SWRNHSIPLET KDGRQVPOSCRRYRLATIANFSAMGLEPGRDVLGQLESCLDGWEYD 120
QY 121 QDYVLSVTWTEWNLVCEENKWKPLTTSLSFFVGVLLGSFVSGQLSDREGRKNVLFATMAVQ 180
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 KDYFLSTIVTEWNLVCEENKWKPLTTSLSFFVGVLLGSFVSGQLSDREGRKNVLFATMAVQ 180
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 TGFSLQIFISWEMFTVLFVIVGMQISNYVAVFTLGTTEILGKSVRIIFSTLGVCTFFA 240
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 TGFSLQIFISWEMFTVLFVIVGMQISNYVAVFTLGTTEILGKSVRIIFSTLGVCTFFA 240
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 241 VGYMLPLFAFYFIRDRWMLLALATVPGVLCVPLWMTIPESPRWLISQRRFREAEIIOKA 300
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 IGYMVLPLFAFYFIRDRWMLLALATVPGVLCVPLWMTIPESPRWLISQRRFREAEIIOKA 300
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 AKMNTAVPAVIEDSVF--ELNPLKQOKAFILDLFTRNIAIMTMSLLWMLTSVGYFA 358
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 301 AKMNTAVPAVIEDSVF--ELNPLKQOKAFILDLFTRNIAIMTMSLLWMLTSVGYFA 358
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAAVLFWGGVLLFIQIV 418
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 361 LSLNPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAGVLFWGGVLLFIQIV 420
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 419 PVDYFSLGVLMLGKFGITSAFMSLYVFTAEYPTLVNMAVGVTSASRVGSIITAPYF 478
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 421 PEDYFVSLGVLMLGKFGITSAFMSLYVFTAEYPTLVNMAVGVTSASRVGSIITAPYF 480
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 479 VYLGAYNRMPLPYIVMGSLTVLIGLFTLFPESLGMTLPETLEOMQVKWFRSGKKTDRSM 538
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 481 VYLGAYNRMPLPYIVMGSLTVLIGLFTLFPESLGMTLPETLEOMQVKWFRSGKKTDRSM 538
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 539 ETEENPKVLITAF 551
:||||:||||:
Db 541 DREENPKVLITAF 553
:||||:||||:
RESULT 5
Q9WTN6 PRELIMINARY; PRT; 564 AA.
ID Q9WTN6 AC Q9WTN6;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE ORGANIC CATION TRANSPORTER.
GN SLC22A9 OR OCTN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Nezu J.;
RT "Mouse OCTN3 - a novel OCTN transporter family protein.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AB018436; BAA78343.1;
DR MGI; MGI:1929481; SLC22a9.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 564 AA; 63320 MW; C37FDCA6395DAD01 CRC64;
```

```
Query Match 71.5%; Score 2034.5; DB 11; Length 564;
Best Local Similarity 69.0%; Pred. No. 3.3e-117;
Matches 380; Conservative 75; Mismatches 85; Indels 11; Gaps 3;

QY 1 MRDYDEVIAFLGEGWGPFORLIFFLLSASIIIPNGFNGMSVVFVLAGTPHRCRVPDPAANLSS 60
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MRDYDEVIAFLGEGWGPFORLIFFLLSASIIIPNGFNGMSVVFVLAGTPHRCRVPDPAANLSS 60
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
QY 61 AWRNNSVPLRLDRCREVPVHSCSRYLATIANFSALEGRDVLQDLQESCLDGEWFS 120
Db 61 AWRNHSIPMETKQCEPVQKCRRYRLATIANFSELGLEPGRVDLQDLQESCLDGEWYD 120
QY 121 ODVYLSVTVTEWNLVCEDNKKVPLTSLFVGVLLGSFVSGQLSDFRGRKNVLFPATMAVO 180
Db 121 KDFLSTIVTEWDLVCKDDWKAPLTTSFYVGVLLGSFVSGQLSDFRGRKNILFTMAH 180
QY 181 TGFSLQIFISWEMTVLFIIVGMGOISNYVAVFLITGLKSKSVRIIFSTGLVCTFEA 240
Db 181 TGFSTQVFSVNEFMTLLTYLVGMGHISNYVAAFVIGTEMLSKSVRIIFATLGVCCIFFA 240
QY 241 VGYMLPLFAFYTRDRWMLLALTLPVGLVLPWFIPESRWLISQRRFRAEDIQKA 300
Db 241 FGFWVLPFAFYTRWRLLLAITLPGVLCALWVFPESRWLISQGRKEAEVIIRKA 300
QY 301 AKMNTAVPAVIFDSVEELNPL-----KQKAFIDLPFRTRNIAIMTMSLLNMLTSV 354
Db 301 AKINGIVAPSTIFDP-SETNKLQDSSKKPQSHHYDVLRTNIRILTMSIILMTISV 359
QY 355 GYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRXTIAAVLFWGGVLLF 414
Db 360 GYFGLSIDPFLNGLNYVNCFLAAVEPAYVLAWLLQHVSRYSMAGSLFGGVL 419
QY 415 IQLVPVDYFLSTGLVYMLKFGITSFMSLYVFTAEYLTPLVRNMAVGTSTASRVGSII 474
Db 420 VOLVPSDLHLVSTLTMVWGKFGITSAYSVMVYVTAELYPTVVRNMGVGSSTASRLGSTL 479
QY 475 APYFVYLGAYNMLPVIYVNGSLTVLIGITLFFPESLGMTLPTLEQMKQVWFRSGKKT 534
Db 480 SPYFVYLGAYDRRLPYILMGSLTILTAITLFFPESSGVSLPETIDEMQVKKL----KQ 535
QY 535 RDSMETEENPK 545
Db 536 RQSLSKKSPK 546
RESULT 6
Q9VCA2 PRELIMINARY; PRT: 548 AA.
AC Q9VCA2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ORCT PROTEIN.
GN ORCT OR CG6331.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
```

```
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner B.C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AE003747; AAF56271.1;
DR FlyBase; FBgn0019952; Orct.
DR InterPro; IPR003662; sub.transporter.
KW Pfam: PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 548 AA; 61002 MW; 08D7F97599B477AF CRC64;

Query Match 30.1%; Score 857; DB 5; Length 548;
Best Local Similarity 35.7%; Pred. No. 4.8e-45;
Matches 201; Conservative 98; Mismatches 208; Indels 56; Gaps 10;

QY 4 YDEVIAFLGEMGPQRLIFLLSASIIIPNGFMGVVFLAGTPEHRCVP----DAANLS 59
Db 3 YDDVITHLGEFGPYQKRIYVLLCPALVCAFHKLAVGLLAKDFRCALPYENGSIYELS 62
QY 60 SAWENNSVPLRLDRGREVPVHSCSRYLATIANFSALEGRDVL-----COL----- 108
Db 63 PHLNLSYPENER-----CSYI-----DVDYTEEYLNCSIPRSSN 97
QY 109 EQESCLDGEWFSQDVLSTVVTWNLVCEDNKKVPLTSLFFVGVLLGSFVSGQLSDFRG 168
Db 98 ETKTC-SSYVVDKSKYLNSAVTEWNLVCSRLSLSATSDFMLGVLLGLSFLIFQMSDKLG 156
QY 169 RKNVLFATMAVQGFSPLOIFSTISWEMFTVLFVIVGMGOISNYVAVFLITGLKSKSVRI 228
Db 157 RKPTFFASLVLIQIFGVLAAPAPEYFTTSRMIVGATTSGLVAVVIALEMGVSSYR- 215
QY 229 IFSTLGVCTFFAVGYMLLPLFAFYTRDRWMLLALTLPVGLVLPWFIPESRWLISQR 288
Db 216 LFAGVAMQMFSSVGFMLTAGFAYFIDWRWLQITAILPGLLFLCYWIIIPESARWLLMKG 275
QY 289 RFRRAEDIIQAAKAMNNTAVPAVIFDS-VEELNPLKQQ-----KAFILDLFRTNRNIA 339
Db 276 RKDEAFVIEKAARENKVEPNEIEQLVDEVAEKKKQDEMAASQPAATVFDLLRYPNLR 335
QY 340 IMTMSLLNMLTSVGYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRY 399
Db 336 RKTLLIFDFNFVNGVYIGLSNNTNLGGNQLNFMISGAVEIPGYVTLTLTLNRRGRS 395
QY 400 ITAAVLFWGGVLLFQLVDPVDYFELSIGLVMLKFGITSFMSLYVFTAEYLTPLVRN 459
Db 396 ILGCTMVAAGISLLATIFVPSDMNMLIVACAMIGKLAITSYGTITFSAEQPTVVRNV 455
QY 460 AVGTSTASRVGSIIAPFYFYLGAYNMLPVIYVNGSLTVLIGITLFFPESLGMTLPTL 519
Db 456 GLGASSMVARVGGILAPYLKLLGEIKWRPLIICGALSITAGLLSLLPTELNKPMPETI 515
QY 520 EQMKQVWFRSGKKTDRSDMETEE 542
Db 516 EDGENF-----GKKPAQETAEE 533
```



```

RESULT 7
O01384
ID O01384 PRELIMINARY; PRT; 548 AA.
AC O01384;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE ORGANIC CATION TRANSPORTER.
GN ORCT OR CG6331.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=98072431; PubMed=9409773;
RX Taylor C.A.M., Stanley K., Shirras A.D.;
RT cation transporter with six or 12 transmembrane domains.*;
RL Gene 201:69-74(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; Y12400; CAA73031.1;
DR FlyBase; FBgn019952; Orct.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR Transmembrane.
KW TRANSMEMBRANE.
SQ SEQUENCE 548 AA; 61105 MW; 2665B71C969C5E4A CRC64;

Query Match 29.9%; Score 852; DB 5; Length 548;
Best Local Similarity 35.5%; Pred. No. 9.8e-45;
Matches 200; Conservative 99; Mismatches 208; Indels 56; Gaps 10;

QY 4 YDEVIAGFGEWGPORLIFELLSSAIIPIGNGMSVVFVLAGTPEHRCRVP-----DAANLS 59
DB 3 YDDVITTHGEFGPKQRIYLLCLPAIVCAFHLKAGVFLAKDFDCAIPYENGSIYELS 62
QY 60 SAWRNNSVPLRLRDREVPCHSCRYRLATIANFSALGLEPGROVDL-----COL----- 108
DB 63 PHLNLSYPENER-----CSYV-----DVDYTEEYLNIGSIPRSSN 97
QY 109 EQESCLDGNFESQDYLSTVVTENLVCEDNWKVPLTTSIFFVGVLLGSFVSQLSDFEG 168
DB 98 ETKTC-SSYVYDRSKYLSNAVTEWNLVCSRLSATSDSLFVLGVLGSLIFGMSDKLG 156
QY 169 RKNVLFATMAVOTGFSFLQIFSIWEMFTLVFIVGMQISNVYVAFILGTEILGKSVRI 228
DB 157 RKPTFFASVLQILFGVLAAVAPEFSYTSIRMLVGATTSGVFLVAVIALEWVGSSYR- 215
QY 229 IFSTLGVCTFFAVGMMLPLFAYFIRDMRLLLALTVGVLCVPLMWFIPESPRLISQR 288
DB 216 LFAGVAMQMFESVGFMLTAGFAYFIHWRMLQIAITLPLGLCYIWIIPESARLLMKG 275
QY 289 RPRAEETIQAAKNNTAPAVIFDS-VVELNPLKQ-----KAFILDIFRTENIA 339
DB 276 RKDEAFVIEKAARENKVEVPEIYEQVLVDEVAEKKKODEMAASQAATVFDLLRVPNLR 335
QY 340 IMTINSLLLWMLTSGVYFALSPLDPAHGDALNCFLSALIEIPAYITAMLLRLTPRRY 399
DB 336 RTLLIFDFWNGSVYGLSNWNTNNGGNOLVNFMLSGVPEIPGYITLLFFTLNRWGRS 395
QY 400 ITAAVLEWGGVGLFLOLVDPYDYFLSTIGLVMLGKFGITSAFSMLYVFTAEYPTVIRNM 459
DB 396 ILCGTMWVAGISLLATIFVPSDMNWLIVACAMIGKLAITSYGTIYIFSAEQEPTVVRNV 455
QY 460 AVGVTSTASRGSIIAPYFVYLGAYNMPLPYIVMGSLTVLIGITLFFPESLGMTLPETL 519
DB 456 GIGASSMVARVGGIAPYKLLGLTWRLPLIICGALSILTAGLLSRLLPETLNKPMPTI 515
```

```

QY 520 EQMKVKWFRSGKKTRDSMETEE 542
DB 516 EDGENF-----GKAPAQETAEE 533

RESULT 8
O9VCA3
ID O9VCA3 PRELIMINARY; PRT; 567 AA.
AC O9VCA3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CGL3610 PROTEIN.
GN CGL3610.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weisbrock G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AB003747; AAF56270.1;
DR FlyBase; FBgn0039176; CGL3610.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR Transmembrane.
SQ SEQUENCE 567 AA; 63166 MW; 0D9B979469395E60 CRC64;

Query Match 28.8%; Score 819; DB 5; Length 567;
Best Local Similarity 34.0%; Pred. No. 1.1e-42;
```



QY	474	IAPYFVVLGAYNRMLPYIVMGSLTIVLIGITFTLFFPESLGMTLPTELEQMKVKWFRSCKK	533
Db	475	LAPHIVNLGKIYKILLPGLMGLMALSAGILTFELPETLGAPLPMTIEDAENP-----GKK	529
QY	534	-----TRDSMETEENP	544
Db	530	PEPDSGMFTQAARKRESQP	548
RESULT	11		
ID	002270		
ID	002270	PRELIMINARY;	PRT; 576 AA.
AC	002270;		
DT	01-JUL-1997	(TrEMBLrel. 04, Created)	
DT	01-JUL-1997	(TrEMBLrel. 04, Last sequence update)	
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)	
DE		ORGANIC CATION TRANSPORTER 1.	
GN	OCT-1.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99227113; PubMed=10209228;		
RA	Wu X., Fei Y.J., Huang W., Chancy C., Leibach F.H., Ganapathy V.;		
RT	"Identity of the F52F12.1 gene product in Caenorhabditis elegans as an		
RT	organic cation transporter."		
RL	Biochim. Biophys. Acta 1418:239-244(1999).		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).		
CC	-1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.		
DR	EMBL; AF110415; AAF21932.1; "		
DR	InterPro; IPR003662; sub-transporter.		
DR	Pfam; PF00083; sugar_tr; 1.		
KW	Transmembrane.		
SQ	SEQUENCE 568 AA; 63505 MW; COA3E73851F44056 CRC64;		
Query Match	26.5%;	Score 754.5;	DB 5; Length 568;
Best Local Similarity	30.2%;	Pred. No. 9.7e-39;	
Matches	169;	Conservative 119;	Mismatches 236; Indels 35; Gaps 9;
QY	3	DYDEVIATFGEWGPFORLIFFLLS-ASIIPIGNFGNMGSVVFLAGTPEHRCRVPDPAANLSSA	61
Db	8	DFDFVLEQVNGYTIQIVFFIICLTSLPSAFSAFNIPFVVGPNPHTCHIPEGKEYLRP	67
QY	62	WRNSVPLRLDRGVRPHSCRSRYRLATIANFSALGLEPDRDVLGQLESCLDGWEFSQ	121
Db	68	LTNDTQIL-----SCKQYNETQINVFRFTSAP-VDYTSRISLVPCQNGWDYDN	116
QY	122	DVYLSTVVTENLVCEDNWKVPLTTSIFVGVLLGSFVSGQLSDRGKRNKLVFATMAVQT	181
Db	117	STYLDLSLVTEFNLCVDOQAWIEISTTSFYVGSFIGNCLGYADKFGRRRSFVILTILI	176
QY	182	GFSPQLQIFSISWEMFTVLFVIVGMQISNVYVAFILGTEILGKSVRIIFSTLGVCTFFAV	241
Db	177	VCGTASSFADIESFIILRFTTGLARPALFOIPFIICMEFMGNSGR-IFSGMLTSLFFGA	235
QY	242	GYMLLPLFAVFIKRWMLLALVPGVLCVPLWVFPESPRNLISQRRPREAEDIQKAA	301
Db	236	AMALLGVVAMFIRWRQLTFFCNAPAFYIIYFFFPESPRNSVSGVKWADARKQKUKIA	295
QY	302	KWN---NTAPAVIFDSVEELNPLKQKAF-----ILDFPRTNIAIMTMSLLWLMTS	353
Db	296	KMKGKSNVDVDELV-DSMKHQNQAABEKWKRSNVTVDLEKTNLRKKTILVIYIWMNA	354
QY	354	VGYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAWLLLTLPRIYIIAAVLFWGGGVLL	413
Db	355	IYINGLTNLSNLPVDDYWSFIINGAVELPFGYFVWVPLQACAGRRWPLAATMIVCGTGV	414
QY	414	FIQLVPDYFELSLGLVNLGKFGITSAFSMLYVETAEIYPTLVNMAVGTSTASRVGSI	473
Db	415	SAMFMPDGPVPLVASFIFGKFGVGSFVYIIFAGELYPTVVRIGMGSSVAGSGLL	474



QY	174	FATMAVOTGFSFLQIFSISWEMFTLVFIVMGQISNYVVAFILGTEILGKSVRIIFSTL	233
		I : : I : : I : : I : : I : : I : : I : : I : : I : :	
Db	181	LTTVLINAAGVLMALSPITYTWMLIFRLIQGLGVSKAGWLIGVILITEFVGRRYR--RTV	237
QY	234	GVC--TFFAVGYMLLPLFAFYFRDRNMLLATATPCVLCVLPVWFIPESPRWLISORRER	291
		I : : I : : I : : I : : I : : I : : I : : I : : I : :	
Db	238	GIFYQVAYTVGLDLVLAGVAYALPHWRWLOFTVALPNFFELLYWCIPSPRWLISQNKA	297
QY	292	EAEEDITQAKAMNNTAVPAVI-----FDVSVEELNP LKQKAFILDLFTRNIAlTIMS	345
		I : : I : : I : : I : : I : : I : : I : : I : : I : :	
Db	298	EAMRIIKHIAKKNGKSPLASQLRLEEEETGGKLN-----SFLDLVPTQIRKHTMIL	351
QY	346	LLLWMLTISGVGFALSADPNLHCD-AYLNCFLSALLIEIPAYITAWMLLTLPTRYIIAAV	404
		I : : I : : I : : I : : I : : I : : I : : I : : I : :	
Db	352	MYWNFTSSVYLQGLIMHM-GLAGDNITLDFYSALVEFFAAPFMILTIDRGRRYPWAAS	410
QY	405	LFWGCGVLLFIQLVPVDYYFLSIGLVMLGKFGITSFMSLYVFETAELYPTLVRRNAVGV	464
		I : : I : : I : : I : : I : : I : : I : : I : : I : :	
Db	411	NMVAGAACLASFVPGDLQWLKIIISCLGRMGITMAYEIVCLVNLAELYPTFIRNLGVHC	470
QY	465	STASRVGSIIAPYVILGYANRM-----LPYIVMSGSLTVLIGIFTLFPESLGMTLPTT	518
		I : : I : : I : : I : : I : : I : : I : : I : : I : :	
Db	471	SSMCDGGIITPPLVY-----RLTNWLELPLMFVGVGLVAGGLVLLLPETKGALPET	525
QY	519	LEQMOKVKWPRSCKK 533	
		I : : I : : I : : I : : I : : I : : I : : I : : I : :	
Db	526	IEEAENMQRPKRKNKE 540	
		I : : I : : I : : I : : I : : I : : I : : I : : I : :	
 RESULT 14 002713 PRELIMINARY; PRT; 554 AA.			
ID	002713	PRELIMINARY;	PRT; 554 AA.
AC	002713		
DT	01-JUL-1997 (TREMBLrel. 04, Created)		
DT	01-JUN-1997 (TREMBLrel. 04, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	APICAL OPTICAL CATION TRANSPORTER.		
GN	OCT-2P.		
OS	Sus scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
RX	NCBI_TaxId=9823;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	TISSUE=KIDNEY;		
RX	MEDLINE=97256752; PubMed=9099681;		
RA	Gruendemann D., Babin-Ebell J., Martel F., Oerding N., Schmidt A.,		
RA	Schoenig E.;		
RT	"Primary structure and functional expression of the apical organic		
RT	cation transporter from kidney epithelial LLC-Pk1 cells.";		
RL	J. Biol. Chem. 272:10408-10413(1997).		
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).		
CC	-!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.		
DR	EWBL; Y09400; CAA70567.1;		
DR	InterPro: IPR003662; sub_transporter.		
DR	Pfam: PF00083; sugar_tr; 1.		
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.		
KW	Transmembrane.		
SW	SEQUENCE 554 AA; 61989 MW; E14B556560C553B CRC64;		

```

Query Match      24.6%; Score 700; DB 6; Length 554;
Best Local Similarity 34.0%; pred. No. 2.1e-35;
Matches 188; Conservative 94; Mismatches 237; Indels 34; Gaps 13;

Qy 1 MRDDEVIAFTGEGWGPFORLIFF---LLSASIIIPNGMGMSVWFLAGTPEHRCRVPDAA 57
      | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MLTVDDILEHTGERNFQKFQFFLLALLSAAFTP---IYVGIVFLGFTPDHRCRSPGVAE 57

Qy 58 LSS-----NNSVPLRLDRGVEPHSCSRVRLATIANFSAIG--LEP--GRDVLGQ 107
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 58 LSORCGWSIAPELNTYVPGPGPAGAAFFPCRRCRRYEDVM--NOSTLGGCVDPDLGLAANSSH 115

```

QY	108	LEBESCLDGHFESQDYLSTVTWNVLNVEDNWKVPLDTTSLFVGVLLGSFVSGQSLDSRE	167
Db	116	LPLGPRCYGWY--DTPGSSIVTEFDLVCAVNSLLDFQSANVNGFFIGSVGIGYIADRE	173
QY	168	GRKNVLFATNAVQTGFSLQIFISISWEMFTVLFVIVGMGOISNVYVAFILGTEILGKSVR	227
Db	174	GRKCLLLTLLINAVSGVLMASPTYTWMVLPRLIOGLVSKAGWMIGYILITEFVGLSYR	233
QY	228	IIFSTLGVCTTFPAGVYMLLPFAFYFIRDMRLMLLALIVPGVLCVPLWVFPESPRWLI	285
Db	234	--RTVGIFQVAFTFGLLVAGVAYALPHWRVLOFTVLPNFCFFLYWCVPESPRWLI	290
QY	286	SORFREAEEDIOKAKMNNTPAVIFDSVSELNPKO---OKAFILDLERTRIAM	341
Db	291	SONKAKAMSLIKHIAKNGKSIPA---SLQSLRDEEVEGKLFPSLDLVYTPQIRKH	346
QY	342	TIMSLLLMLTSGVGFALSDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLPLRRII	401
Db	347	TLILMYNWFTSAVLYQGLVHMGLAGSLNLYLDFYSALVEFFPAALLTLITIDLGRHPW	406
QY	402	AAVLFWGGVLLFTQLVPDYVYFLSICGLVMGLKFGTISAFSMLYVFTAELYPTLVNMAV	461
Db	407	AASNWAGAACLVFIPEDPHMLRITVLCIGRMGITMAYEVCVNNAEIYPIFINLGV	466
QY	462	GVTSTASRSGSIAPYEVY-IGAYNRMPLVIVMGSLTILIGITFLTFEPFSGMGLTPETIE	520
Db	467	LVCSSMCDIGIITPFLVYRLTDIWHELPLVFAVAVGLIAGGLVLLLPETKGTLPETIE	536
QY	521	QMKVKWFRSGKK	533
Db	527	EAETMRPRPKNK	539
RESULT 15			
P70485			
ID	P70485	PRELIMINARY;	PRT; 593 AA.
AC	P70485;		
DT	01-FEB-1997 (TrEMBLrel. 02, Created)		
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	ORGANIC CATION TRANSPORTER OCT2.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;		
NCBI_Taxid=10116;			
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-SPRAGUE-DAWLEY; TISSUE=KIDNEY;		
RX	MEDLINE=9629517; PubMed=8702418;		
RA	Okuda M., Saito H., Urakami Y., Takano M., Inui K.;		
RT	"cDNA cloning and functional expression of a novel rat kidney organoid		
RT	cation transporter, OCT2."		
RL	Biochem. Biophys. Res. Commun. 224:500-507(1996).		
CC	!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).		
CC	!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.		
DR	EMBL; D83044; BAAL1754.1;		
DR	InterPro; IPR003662; sub_transporter.		
DR	Pfam; PF00083; sugar_tr; 1.		
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.		
KW	Transmembrane.		
SO	SEQUENCE 593 AA; 66080 MW; 36C1044E0C04B443 CRC64;		

```

Query Match      24.6%; Score 699.5; DB 11; Length 593;
Best Local Similarity 34.0%; Pred. N.2.4e-35;
Matches .189; Conservative 87; Mismatches 243; Indels 37; Gaps 14;

y      1 MRDYDEVIAFLGEGWGFQRLIEFLLSASTIPNGENG--MSVVFLAGTPEHRCRVPDAANL 58
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
b      1 MSTVDDILEHGEHFHUFQKOTFLL--ALLSGAFPTYYGVIGVFLGTDPDHWCSPGAKL 58
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

y      59 SS--AMR-----NNSVP-LRLRDGREVPHSCSYR-----LATIANFSAULGLEPGRDV 103
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 59 SORCWSQAELNYTPVGLPSDEASFLSQCMRYEVDWQSTLDCVDPSSLA-----A 112  
QY 104 DLGLOEQESCLDGWFSQDYLSTVTEWNLVCEDNKYPPLTSLFFVGVLLGSFVSGOL 163  
Db 113 DRNQLPLGCEHGWYNTFG--SSIVTEFNLCAHSMIDLQFQSVVNVGFFIGAMMIGYL 170  
QY 164 SDRFGRKNVLFATPMAYOTGFSFLQIFSIWEMFTVLFIWGMGOISNYVVAFILGTEILG 223  
Db 171 ADRFGRKFCLLVTILINAISGALMAISPNYANMLVFRFLQGLVSKAGWLGILITEFVG 230  
QY 224 KSVRIIFSLGVC-LTFVAVGYMLPLPAYFIRDNRMLLALLTVPGLVCPVLWFIPEBP 281  
Db 231 LGYR---RMVGICYAFTVGLLILAGVAYIPNWRWLOFAVTLFNFCEFLYFNCIPESP 287  
QY 282 RWLISQRRFREADIIQKAAKNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNIAIM 341  
Db 288 RWLISQNKIVKAMKIKHAKKNGKSVPSLQNLTPDEDAKGLKPSILDLVRTPOIRKH 347  
QY 342 TMSLLMLMTSVGYFALSADAPNLHGD-AYLNCFLSALIEIPAYITAWLLRLTLPRRYI 400  
Db 348 TLILMYNFTSSVLYOGLJMHM-GLAGDNIYLDFFYSALVEPPAAFIILITIDRVGRYP 406  
QY 401 IAAVLFWGGGVLLFQLVPVDYFFLSIGLVMLCKFGITSAFSMLYVFTAELYPTLVNRMA 460  
Db 407 WAVSNVAGAACLASVFIPTDDLOWLKITTACLGMRGITMAYEMVCLVNAELYPTVIRNLG 466  
QY 461 VGYTSTASRVGSIIAPFYFY-LGAYNRMLPYIVMGSLTVLIGIFTLPFPESIGMTLPETL 519  
Db 467 VLVCSMCDIGGIIIPFLVYRLTDIWMEEPLVVFVAVGLVAGALVLLLPETKGKALPETI 526  
QY 520 EQMKVKWFRSGKKTR 535  
Db 527 EDAENMQ--RPRKKER 540

Search completed: January 25, 2002, 10:14:54  
Job time: 335 sec

...

GenCore version 4.5  
 Copyright (c) 1993 - 2000 Compugen Ltd.  
 OM protein - protein search, using sw model  
 Run on: January 25, 2002, 10:09:12 ; Search time 43.11 Seconds  
 (without alignments)  
 957.059 Million cell updates/sec

Title: US-09-521-195-3  
 Perfect score: 2883  
 Sequence: 1 MRDYDEYTAFLGFWPQRL.....HTRMLKDGQERPIILKSTAF 557

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
 Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
 1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
 2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
 3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
 4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
 5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
 6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
 7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
 8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
 9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
 10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
 11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
 12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
 13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
 14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
 15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
 16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
 17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
 18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
 19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
 20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
 21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
 22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2883	100.0	557	20 AAY01650	A protein with cat
2	2883	100.0	557	21 AAY83929	Human carnitine tr
3	2513	87.2	557	20 AAY01652	A protein with cat
4	2513	87.2	557	21 AAB20360	Mouse OCTN2 amino
5	2513	87.2	557	21 AAY83930	Mouse carnitine tr
6	2295.5	79.6	564	21 AAB20578	Mouse OCTN3 protei
7	2218	76.9	551	20 AAY01649	A protein with cat
8	2148	74.5	553	20 AAY01651	A protein with cat
9	2148	74.5	553	21 AAB20579	Mouse OCTN1 amino
10	785.5	27.2	560	21 AAB34308	Human ORFX ORF2802
11	784	27.2	584	22 AAM00930	Human bone marrow

12	735	25.5	554	19 AAW64538	Human liver cell c
13	724.5	25.1	540	22 AAB49401	Murine organic ani
14	719	24.9	535	21 AAY51249	Rat liver anion tr
15	712	24.7	556	17 AAR77676	Rat OCT-1 protein.
16	701.5	24.3	538	22 AAB47273	HOAT2B. Homo sapi
17	697	24.2	546	22 AAB47272	Human cerebral org
18	693	24.0	542	21 AAY92902	A human organic an
19	690	23.9	548	21 AAB08823	HOAT3.. Homo sapie
20	689	23.9	542	22 AAB47274	A human organic an
21	686.5	23.8	551	21 AAB08824	Human osteoclast t
22	673	23.3	561	18 AAB44196	Human protein havi
23	668	23.2	578	22 AAE06571	Hydrophobic domain
24	660.5	22.9	607	21 AAB12131	Rat cerebral organ
25	656.5	22.8	536	21 AAY92903	Mouse osteoclast t
26	652.5	22.6	537	18 AAW44195	Mouse organic anio
27	644	22.3	545	22 AAB36553	Human organic anio
28	641	22.2	550	21 AAY44278	Human organic anio
29	641	22.2	550	22 AAB47271	HOAT1.. Homo sapie
30	638.5	22.1	551	20 AAW88488	Human organic anio
31	628	21.8	563	20 AAW88489	Human bone marrow
32	621.5	21.6	483	22 AAM00982	Human organic havi
33	607	21.1	550	22 AAE06612	HOAT5. Homo sapie
34	607	21.1	550	22 AAB69091	Human transporter
35	541	18.8	541	22 AAB47276	HOAT4. Homo sapie
36	477.5	16.6	553	22 AAE04897	Arabidopsis thalia
37	451	15.6	554	22 AAB47275	Arabidopsis thalia
38	433	15.0	480	21 AAG28044	Arabidopsis thalia
39	433	15.0	483	21 AAG28043	Arabidopsis thalia
40	380.5	13.2	515	21 AAG42522	Arabidopsis thalia
41	368.5	12.8	521	21 AAG30885	Human organic cati
42	357.5	12.4	548	21 AAY44633	Arabidopsis thalia
43	344.5	11.9	397	21 AAG30886	Corynebacterium gl
44	332	11.5	439	22 AAB76766	C glutamicum prote
45	332	11.5	448	22 AAG93041	

## ALIGNMENTS

RESULT 1  
 AAY01650  
 ID AAY01650 standard; Protein; 557 AA.  
 AC AAY01650;  
 XX  
 XX 23-JUN-1999 (first entry)  
 DT  
 XX A protein with cation transporting activity.  
 DE Organic cation transporter; OCT1; OCT2; drug development; fatty liver;  
 KW heart disease; cancer; anti-tumour drug; anticancer drug.  
 OS Homo sapiens.  
 XX  
 XX WO9913072-Al.  
 PN  
 XX  
 PD 18-MAR-1999.  
 XX  
 PF 07-SEP-1998; 98WO-JP04009.  
 XX  
 PR 20-MAY-1998; 98JP-0156660.  
 PR 08-SEP-1997; 97JP-0260972.  
 XX  
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 PA Nezu J, Oku A;  
 PI WPI; 1999-215062/18.  
 DR N-PSDB; AAX26880.  
 DR  
 XX Genes homologous with organic cation transporters OCT1 and OCT2,  
 PT useful in design of new drugs for treatment of diseases due to  
 PT abnormality of the transporter functions

```
XX PS Claim 1; Page 51-55; 97pp; Japanese.
XX CC The present sequence represents a protein with cation transporting
CC activity. The genes are significantly homologous with organic cation
CC transporters OCT1 and OCT2. The genes may be used in drug development,
CC particularly in the treatment of diseases due to abnormality of the
CC organic cation transporter functions e.g. fatty liver, heart diseases
CC and cancers, by controlling such as by inhibition or activation.
CC Administration of anti-tumour and anticancer drugs in combination with
CC a transporter protein inhibiting agent allows the agents to penetrate
CC into the diseased cells to enhance the drug action.
XX SQ Sequence 557 AA;

Query Match 100.0%; Score 2883; DB 20; Length 557;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEGWGPQRLIFLLSASITPNFTGLSSVFLIATPEHRCRVPDAAANLSS 60
DB 1 mrdydevtaflgewgpqrlifllsasiipngftglssvfliatpehrcrvpdaanlss 60

QY 61 AWRNHTVPLRLDRGVRPHSCRRYRLATIANFSALEPGRDVDLQLEQESCLDGWEFS 120
DB 61 awrnhtvplrlldrgrevphscrryrlatiansalglepgrdvdldgleqescldgwefs 120

QY 121 QDYLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGRKNVLFVTMGQ 180
DB 121 qdylstivtewnlvceddwkaptislsffvgvllgsfsgqlsdrfgrknvlfvtmgmq 180

QY 181 TGFSELIQIFSKNFEMFVFLVGVGMGOISNYAAFLVGLTEILGKSVRIIFSTLGVCFYA 240
DB 181 tgfseqliqifsknfemfvflvvgmgqisnyaaflvlgteilgksvriifstlvgvcifya 240

QY 241 FGVMVPLPFAYPTRDWRMLLVALTPGVLCAVLMWFIPESPRWLISQGRFEEAEVIRKA 300
DB 241 fgvmvplpfayptrdwrmlvaltmgpvlcavlmwfiipesprwlisqgrfeaeaeiirka 300

QY 301 AKANGIVPSTIFDPSLQDLSSKKQOSHNIIDLLRTNIRMTIMSLMWTISVGYFG 360
DB 301 akangivvpstifdpselqdlsskkqoshniidllrtwnirmvtimsmlmwtisvgyfg 360

QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLQLLOYPRLRYSMATLFLGGSVLLFMQLV 420
DB 361 lsldtpnlhgdifvncflsamvevpayvlawlqlloyprrysmatalfggsvllfmqlv 420

QY 421 PPDLYYLATVLMVGMKFGVTAAFSMVYVYTAELPTVTVVRNMGVSVSTASRLGSLSPYF 480
DB 421 ppdlYYLATVLMVGMKFGVTAAFSMVYVYTAELPTVTVVRNMGVSVSTASRLGSLSPYF 480

QY 481 VYLGAYDRFLPYLLMGSLTILTAITLFLPESFGPPLDITDQMLRVKGMKHKRTPSHTR 540
DB 481 vylgaydrflpyllmgsltiltailtflpesfgppldtidqmlrvkgmkhkrtpshtr 540

RESULT 2
AA183929
ID AAY83929 standard; protein; 557 AA.
XX AC AAY83929;
XX DT 05-JUL-2000 (first entry)
XX DE Human carnitine transporter protein OCTN2.
XX KW Organic cation transportation; human; carnitine transporter protein;
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
```

```
KW OS juvenile visceral steatosis.
XX OS Homo sapiens.
XX PN W0200014210-A1.
XX PD 16-MAR-2000.
XX PF 07-SEP-1999; 99WO-JP04853.
XX PR 07-SEP-1998; 98JP-0252683.
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX PI Nezu J, Oku A;
XX WIPI; 2000-256966/22.
XX DR N-PSDB; AAA09889.
XX PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
XX cation transporter, useful as diagnostic tool -
XX PS Example 1; Page 36-40; 106pp; Japanese.
XX CC This sequence represents the human carnitine transporter protein OCTN2.
XX CC The coding sequence can be used as a target for diagnosis of systemic
XX CC carnitine deficiency by detecting the presence of mutations in the
XX CC sequence, especially seen in the disease juvenile visceral steatosis
XX CC (jvs). The wild type OCTN2 gene can be used in the gene therapy of the
XX CC disease state.
XX SQ Sequence 557 AA;

Query Match 100.0%; Score 2883; DB 21; Length 557;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEGWGPQRLIFLLSASITPNFTGLSSVFLIATPEHRCRVPDAAANLSS 60
DB 1 mrdydevtaflgewgpqrlifllsasiipngftglssvfliatpehrcrvpdaanlss 60

QY 61 AWRNHTVPLRLDRGVRPHSCRRYRLATIANFSALEPGRDVDLQLEQESCLDGWEFS 120
DB 61 awrnhtvplrlldrgrevphscrryrlatiansalglepgrdvdldgleqescldgwefs 120

QY 121 QDYLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGRKNVLFVTMGQ 180
DB 121 qdylstivtewnlvceddwkaptislsffvgvllgsfsgqlsdrfgrknvlfvtmgmq 180

QY 181 TGFSELIQIFSKNFEMFVFLVGVGMGOISNYAAFLVGLTEILGKSVRIIFSTLGVCFYA 240
DB 181 tgfseqliqifsknfemfvflvvgmgqisnyaaflvlgteilgksvriifstlvgvcifya 240

QY 241 FGVMVPLPFAYPTRDWRMLLVALTPGVLCAVLMWFIPESPRWLISQGRFEEAEVIRKA 300
DB 241 fgvmvplpfayptrdwrmlvaltmgpvlcavlmwfiipesprwlisqgrfeaeaeiirka 300

QY 301 AKANGIVPSTIFDPSLQDLSSKKQOSHNIIDLLRTNIRMTIMSLMWTISVGYFG 360
DB 301 akangivvpstifdpselqdlsskkqoshniidllrtwnirmvtimsmlmwtisvgyfg 360

QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLQLLOYPRLRYSMATLFLGGSVLLFMQLV 420
DB 361 lsldtpnlhgdifvncflsamvevpayvlawlqlloyprrysmatalfggsvllfmqlv 420

QY 421 PPDLYYLATVLMVGMKFGVTAAFSMVYVYTAELPTVTVVRNMGVSVSTASRLGSLSPYF 480
DB 421 ppdlYYLATVLMVGMKFGVTAAFSMVYVYTAELPTVTVVRNMGVSVSTASRLGSLSPYF 480

QY 481 VYLGAYDRFLPYLLMGSLTILTAITLFLPESFGPPLDITDQMLRVKGMKHKRTPSHTR 540
DB 481 vylgaydrflpyllmgsltiltailtflpesfgppldtidqmlrvkgmkhkrtpshtr 540
```



QY 541 MLKDGQERPTILKSTAF 557  
 Db 541 mlkdqgerptilkstaf 557  
 RESULT 3  
 AAY01652  
 ID AAY01652 standard; Protein; 557 AA.  
 AC AAY01652;  
 XX  
 DT 23-JUN-1999 (first entry)  
 XX  
 DE A protein with cation transporting activity.  
 XX  
 KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;  
 KW heart disease; cancer; anti-tumour drug; anticancer drug.  
 XX  
 OS Mus musculus.  
 XX  
 PN W09913072-A1.  
 XX  
 PD 18-MAR-1999.  
 XX  
 PF 07-SEP-1998; 98WO-JP04009.  
 XX  
 PR 20-MAY-1998; 98JP-0156660.  
 PR 08-SEP-1997; 97JP-0260972.  
 XX  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX  
 PI Nezu J, Oku A;  
 XX  
 DR WPI; 1999-215062/18.  
 DR N-PSDB; AAX26902.  
 XX  
 PT Genes homologous with organic cation transporters OCT1 and OCT2,  
 PT useful in design of new drugs for treatment of diseases due to  
 PT abnormality of the transporter functions  
 XX  
 PS Claim 1; Page 75-79; 97pp; Japanese.  
 XX

CC The present sequence represents a protein with cation transporting  
 CC activity. The genes are significantly homologous with organic cation  
 CC transporters OCT1 and OCT2. The genes may used in drug development,  
 CC particularly in the treatment of diseases due to abnormality of the  
 CC organic cation transporter functions e.g. fatty liver, heart diseases  
 CC and cancers, by controlling such as by inhibition or activation.  
 CC Administration of anti-tumour and anticancer drugs in combination with  
 CC a transporter protein inhibiting agent allows the agents to penetrate  
 CC into the diseased cells to enhance the drug action.  
 XX  
 SQ Sequence 557 AA;

Query Match 87.28; Score 2513; DB 20; Length 557;  
 Best Local Similarity 85.5%; Pred. No. 2.5e-265;  
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEGWGPORLIFLILLSAIIIPNGFTGLSSVFLIATPESHRCRVDPDAANLSS 60  
 Db 1 mrdydevtaflgwgppqrliflilsaipngfngmsivflactpchrclvphbtniss 60  
 QY 61 AWRNHTVPLRLDRGVEPHSRFLATIANFSAIGLEPGRDVDLGOLEQESCLDGEWFS 120  
 Db 61 awrnhsipletkdrqvpqkrrylatianselglepgrdvldleqescldgweyd 120  
 QY 121 QDVYLSTIVTWNLCEDDKWAPITISLFFVGVLLGSFISGQLSDRGRKNVLFVTMGMQ 180  
 Db 121 kdvlstivtewdlvckddkwapititslffvgvlmgsgfisdgrknvlfvtmgmq 180  
 QY 181 TCFSEFLOIFSKNFENFVFLVFLVGMGQISNYAAAFVLGTEILGKSVRIITSLTGLVCIFYA 240

Db 181 tcfseflqfsvnfemftvflvlgmgqisnyaaafvlgteillsksriifatlqvcifya 240  
 QY 241 FGWVLPPLFAYFIRDRMLLVALTMPGVLCVALWFFPESPRWLISQGRFEAEVIRKA 300  
 Db 241 fgwvlpplfayfirdwrmlllaltvpgvlgcalwffipesprwlisqgrfkeaevirka 300  
 QY 301 AKANGIVVPSITFDPSELQDLSSKKQSHNLDLRTWNIRMTIMSTMLWMTISVGVEFG 360  
 Db 301 akingivapstifdpseqlqdlsskqshnldlrtwnirmtimstmlwmtisvsgvfg 360  
 QY 361 LSLDTPNLHGDIFFNCFLSAMVEYPAYVLAWLLQYLPRRYSMATALFLGGSVLLFMOLV 420  
 Db 361 lsldtpnlhgdifvncflsamveypayvllawllqylprrysmatalfggsvllfmqlv 420  
 QY 421 PPDLYLATLVLMVGKFGVTAAFSMAVYVTAELYPTVVRNMGVGVSSASTSLSPYF 480  
 Db 421 pseyfslstalmvgkfgitaysmvyvtaelyptvvrnmgvgvssastslspylpyf 480  
 QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGPLPDTIDQMLRVKGMKHKRTPSHTR 540  
 Db 481 vylgaydrflpyilmgsltiltailtflpsefgyplpdtidqmlrvkgikwqdgqstr 540  
 QY 541 MLKDGQERPTILKSTAF 557  
 Db 541 mlkdqgeesptilkstaf 557  
 RESULT 4  
 AAB20580  
 ID AAB20580 standard; Protein; 557 AA.  
 AC AAB20580;  
 XX  
 DT 11-DEC-2000 (first entry)  
 XX  
 DE Mouse OCTN2 amino acid sequence.  
 XX  
 KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;  
 KW identification; regulator; carnitine transport.  
 XX  
 OS Mus musculus.  
 XX  
 PN W0200046368-A1.  
 XX  
 PD 10-AUG-2000.  
 XX  
 PF 04-FEB-2000; 2000WO-JP00619.  
 XX  
 PR 05-FEB-1999; 99JP-0028406.  
 XX  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX  
 PI Nezu J, Ose A;  
 XX  
 DR WPI; 2000-586982/55.  
 XX  
 PT Organic cation transporter gene OCTN3 expressed in testis for  
 PT identification of regulators of carnitine transport for use as drugs  
 XX  
 PS Example 2; Fig 2; 58pp; Japanese.  
 XX  
 CC The present invention describes a mouse organic cation transporter  
 CC protein (OCTN3). Also described are: (1) a method for screening  
 CC compounds for their ability to regulate the transport of an organic  
 CC cation into the cell, by generating a cell expressing OCTN3 at the  
 CC cell membrane, contacting with the compound and organic cation, and  
 CC observing the degree of transport of the organic cation; and (2) a  
 CC method for screening compounds for their ability to be transported into  
 CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell  
 CC membrane, contacting with the compound and observing the degree of  
 CC transport of the compound. OCTN3 can be used for the identification of  
 CC regulators of the transport of organic cations (especially carnitine)

CC into cells by OCTN3, for use as drugs. The present sequence represents  
 CC the mouse OCTN2 amino acid sequence, which is used in an example from  
 CC the present invention.

XX Sequence 557 AA;

Query Match 87.2%; Score 2513; DB 21; Length 557;  
 Best Local Similarity 85.5%; Pred. No. 2.5e-265;  
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEGWGPQRLIFLFSASIPNGFTGLSSVFLIATPEHRCRVPDAAANLSS 60  
 Db 1 mrdydevtaflgwgppqrlifflsasiipngftglssvfliatpehrcrlvphvtvnls 60  
 QY 61 AWRNHTVPLRLDGRVPHSCRRYRIATANFSALEGRVDVLDGLOESCLDGEWFS 120  
 Db 61 awrnhtvplrlkgrevphscrryriatiansalgelrgrvdlqgleescldgwefs 120  
 QY 121 QDYLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGKKNVLFVTMGQ 180  
 Db 121 kdvlstivtewldvckddwkaptltslffvglmgsfsgqlsdrfgrknvfltmgmq 180  
 QY 181 TGFSLQIFSKNFEMFVLVVGMCQISNYVAAFVLGTEILGKSVRIIFSTLGVCFIYA 240  
 Db 181 tgfsflqvfsvnfemftvlvvgmqisnyvaafvlgteilsksvriifatlvgvcifya 240  
 QY 241 FGVMVLPFLFAYFTRDWRMLLVALTPGVLCLVALWFWIPESRWLISQGRFEAEVILRKA 300  
 Db 241 fgvmvlpflfayftrdwrmlaltvpgvlcgalwfwipesprwlisqgrikeaeavlirka 300  
 QY 301 AKANGIVVPTIFDPSLQDSSKKQOSHNLIDLLRTWIRMTIMSLMWTISVGYFG 360  
 Db 301 akingivapstifdpselqdnstkpqlhhiydlrtwlrmtimslmwtisvgyfg 360  
 QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLOLQVLPDPRYSMATALFLGGSVLLFMQLV 420  
 Db 361 lsldtpnlhgdifvncflsamvevpayvawllolqlvlpdprysmatalfggsvllfmqlv 420  
 QY 421 PPDLXYLATVLMVGKFGVTAAFSMVYVTAELYPTVVRNMGVGSSTASRLGSLSPYF 480  
 Db 421 ppslxylatvlmvgkfgvttaafsmvyvtaelyptvvrnmvgvssstasrlgsilspyf 480  
 QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGPLPDTIDQMLRVKGMKHKRTPSHTR 540  
 Db 481 vylgavdrflpyilmgsltiltaltlflpesfgvplpdtidqmlrvkgikwqqlsqqr 540  
 QY 541 MLKDGQERPTILKSTAF 557  
 Db 541 mqkdgeesptvlkstaf 557

RESULT 5  
 AAY83930  
 ID AAY83930 standard; Protein; 557 AA.  
 XX  
 AC AAY83930;  
 XX  
 XX  
 DT 05-JUL-2000 (first entry)  
 XX  
 DE Mouse carnitine transporter protein OCTN2.  
 XX

KW Organic cation transportation; mouse; carnitine transporter protein;  
 KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;  
 KW juvenile visceral steatosis.

XX Mus musculus.  
 XX WO200014210-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PF 07-SEP-1999; 99WO-JP04853.

XX 07-SEP-1998; 98JP-0252683.  
 XX (CHUG-) CHUGAI, RES INST MOLECULAR MEDICINE INC.  
 XX  
 XX Nezu J, Oku A;  
 XX  
 XX WPI: 2000-256966/22.  
 XX N-PSDB: AAA09890.  
 XX  
 XX Systemic carnitine deficiency gene OCTN2 encoding part of organic  
 XX cation transporter, useful as diagnostic tool -  
 XX  
 XX Example 1; Page 47-51; 106pp; Japanese.

XX This sequence represents the mouse carnitine transporter protein OCTN2.  
 CC The coding sequence of the corresponding human protein can be used as a  
 CC target for diagnosis of systemic carnitine deficiency by detecting the  
 CC presence of mutations in the sequence, especially seen in the disease  
 CC juvenile visceral steatosis (jvs). The wild type OCTN2 gene can be used  
 CC in the gene therapy of the disease state.

XX Sequence 557 AA;

Query Match 87.2%; Score 2513; DB 21; Length 557;  
 Best Local Similarity 85.5%; Pred. No. 2.5e-265;  
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEGWGPQRLIFLFSASIPNGFTGLSSVFLIATPEHRCRVPDAAANLSS 60  
 Db 1 mrdydevtaflgwgppqrlifflsasiipngftglssvfliatpehrcrlvphvtvnls 60  
 QY 61 AWRNHTVPLRLDGRVPHSCRRYRIATANFSALEGRVDVLDGLOESCLDGEWFS 120  
 Db 61 awrnhtvplrlkgrevphscrryriatiansalgelrgrvdlqgleescldgwefs 120  
 QY 121 QDYLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFGKKNVLFVTMGQ 180  
 Db 121 kdvlstivtewldvckddwkaptltslffvglmgsfsgqlsdrfgrknvfltmgmq 180  
 QY 181 TGFSLQIFSKNFEMFVLVVGMCQISNYVAAFVLGTEILGKSVRIIFSTLGVCFIYA 240  
 Db 181 tgfsflqvfsvnfemftvlvvgmqisnyvaafvlgteilsksvriifatlvgvcifya 240  
 QY 241 FGVMVLPFLFAYFTRDWRMLLVALTPGVLCLVALWFWIPESRWLISQGRFEAEVILRKA 300  
 Db 241 fgvmvlpflfayftrdwrmlaltvpgvlcgalwfwipesprwlisqgrikeaeavlirka 300  
 QY 301 AKANGIVVPTIFDPSLQDSSKKQOSHNLIDLLRTWIRMTIMSLMWTISVGYFG 360  
 Db 301 akingivapstifdpselqdnstkpqlhhiydlrtwlrmtimslmwtisvgyfg 360  
 QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLOLQVLPDPRYSMATALFLGGSVLLFMQLV 420  
 Db 361 lsldtpnlhgdifvncflsamvevpayvawllolqlvlpdprysmatalfggsvllfmqlv 420  
 QY 421 PPDLXYLATVLMVGKFGVTAAFSMVYVTAELYPTVVRNMGVGSSTASRLGSLSPYF 480  
 Db 421 ppslxylatvlmvgkfgvttaafsmvyvtaelyptvvrnmvgvssstasrlgsilspyf 480  
 QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGPLPDTIDQMLRVKGMKHKRTPSHTR 540  
 Db 481 vylgavdrflpyilmgsltiltaltlflpesfgvplpdtidqmlrvkgikwqqlsqqr 540  
 QY 541 MLKDGQERPTILKSTAF 557  
 Db 541 mqkdgeesptvlkstaf 557

RESULT 6  
 AAB20578  
 ID AAB20578 standard; Protein; 564 AA.

[illegible]

Qy	301	AKANGIVVPSTIEDPSE---	LQDLSSKKOOSHNIILDLFTWNRIMVVTIMSIMLAWTISVG	357
Dd	301	akingivapstifdpsetnklgdsskkpqshiyolvrtpnrlitmsilwitissvg	360	
Qy	358	yfglsldtPNLHGDI FVNCFLSAMVEPAYVLAWLLLOYLPRRYSMATALEFLGGSVLLFM	417	
Dd	361	yfglsldtpnlngniyvncfilaaevpayvlawlllqhvsrrysmagslfllgsvlllv	420	
Qy	418	QLVPPDLYIYLAIVLMVKRGKVTAAFSMVVYVTAELYPTVVRNMGVGVSVSTASRLGSILS	477	
Dd	421	qlvpsdlhylstclvmvgkfightsaysmvyvtcaelyptcvrnmgvgvsvstasrlgsils	480	
Qy	478	PYFVYLGA YDRFLPYILMGSLLTITALTILFLPESECTPLPDTIDQMLRVKGMKHRTPS	537	
Dd	481	pyfvyलगaydrripyilmgsiltltaitlifpessgvslpetidemqkvxkklkqrqsils	540	
RESULT	7			
AAAY01649	ID	AAAY01649 standard; Protein; 551 AA.		
XX	AC	AAAY01649;		
XX	DT	23-JUN-1999 (first entry)		
DE	DE	A protein with cation transporting activity.		
KW	KW	Organic cation transporter; OCT1; OCT2; drug development; fatty liver heart disease; cancer; anti-tumour drug; anticancer drug.		
OS	OS	Homo sapiens .		
XX	PN	WO9913072-A1.		
XX	PD	18-MAR-1999.		
XX	PF	07-SEP-1998; 98WO-JP04009.		
XX	PP	20-MAY-1998; 98JP-O156660.		
PR	PR	08-SEP-1997; 97JP-O260972.		
XX	PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.		

Dd 1 mrdydeviafgevgpqrllifllsasiipngfngmsvflagtpehrcrvpdaanlss 60  
 Qy 61 AWRNHTVPLRLRDGQREVPHSCRRYRLATIANFSALEPGSDVDLGOLEQESCLDGHEFS 120  
 Dd 61 awrnsvplrlrdgrevphscrryrlatiansaigleprgdvdlgledescldgwefs 120  
 Qy 121 QDVLSTIVTWNLVCEDDWKAPLTISLFFVGLLGSFISQSLSDRFGKKNVLFVFMGMQ 180  
 Dd 121 qdvylstivtwnlvceedwnkvplttslffvglvllgsfsvsgqlsdrfgrknvlfatmavq 180  
 Qy 181 TGFSLQIFSKNFEMFVFLVGVNGQISNYAAVFLGTEILGKSVRIIPSTLGVCIFYA 240  
 Dd 181 tgfslqifsknfemfvlvlgvngqisnyvafilgtelgksvriifstlgvctffa 240  
 Qy 241 FGYNVPLFAYFIRDRMLVALTMPGVLCAVLMWFIPESPRWLISQGRPEAEVIIRKA 300  
 Dd 241 vgymlplfayfirdwrmllaltvpcvlwfwfipesprwlisqrfreacdliqka 300  
 Qy 301 AKANGIVVPSTIFDPSELODSSKKQOSHNLDLRTWNIRMYTMSIMLWMTISVGYFG 360  
 Dd 301 akmntavpavifd--sveelnplkqgkafildlfrtniaimtmllwmltsvgyfa 358  
 Qy 361 LSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLLOVLPYRYSMATALFLCGSVLLFMQLV 420  
 Dd 359 lsldapnlhgdai fncflsalielpayitawllrtpryiaavlfwg99villfiqlv 418  
 Qy 421 PPDIYLIATVLMVGKFGVTAAFSMVYVYTAELYPTVVRNMGVSVSTASRLGSIISPYF 480  
 Dd 419 pvdyyfislglvmlgkfgitsafsmlyvftaelypvltrnmavgtstasrvgsiiapyf 478  
 Qy 481 VYLGAIDRFPLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 540  
 Dd 479 vylgaynrlmipyilmgsltlvltlftlfpesigmlpetleqmkgvkwfrsgk---ktr 535  
 Qy 541 MLKDGQERPTILKSTAF 557  
 Dd 536 dsmeteenpkvl-itaif 551  
 RESULT 8  
 AAY01651  
 ID AAY01651 standard; Protein; 553 AA.  
 AC AAY01651;  
 XX  
 DT 23-JUN-1999 (first entry)  
 XX  
 DE A protein with cation transporting activity.  
 KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;  
 KW heart disease; cancer; anti-tumour drug; anticancer drug.  
 XX  
 OS Mus musculus.  
 XX  
 PN W09913072-A1.  
 XX  
 PD 18-MAR-1999.  
 XX  
 PF 07-SEP-1998; 98WO-JP04009.  
 XX  
 PR 20-MAY-1998; 98JP-0156660.  
 PR 08-SEP-1997; 97JP-0260972.  
 XX  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX  
 PI Nezu J, Oku A;  
 XX  
 DR WPI; 1999-215062/18;  
 DR N-PSDB; AAX26898.  
 XX  
 PT Genes homologous with organic cation transporters OCT1 and OCT2,  
 PT useful in design of new drugs for treatment of diseases due to  
 PT abnormality of the transporter functions

XX  
 PS  
 XX  
 CC The present sequence represents a protein with cation transporting  
 CC activity. The genes are significantly homologous with organic cation  
 CC transporters OCT1 and OCT2. The genes may be used in drug development,  
 CC particularly in the treatment of diseases due to abnormality of the  
 CC organic cation transporter functions e.g. fatty liver, heart diseases  
 CC and cancers, by controlling such as by inhibition or activation.  
 CC Administration of anti-tumour and anticancer drugs in combination with  
 CC a transporter protein inhibiting agent allows the agents to penetrate  
 CC into the diseased cells to enhance the drug action.  
 XX  
 SQ Sequence 553 AA;  
 Query Match 74.5%; Score 2148; DB 20; Length 553;  
 Best Local Similarity 72.2%; Pred. No. 1.8e-225;  
 Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;  
 Qy 1 MRDYDEVIAFGEWGPQRLLIFLLSASIIIPNGFTGLSSVFLIATPEHRCRVDPDAANLSS 60  
 Dd 1 mrdydeviafgevgpqrllifllsasiipngfngmsvflagtpehrcrvpdtvnlls 60  
 Qy 61 AWRNHTVPLRLRDGQREVPHSCRRYRLATIANFSALEPGSDVDLGOLEQESCLDGHEFS 120  
 Dd 61 awrnhsipletkdgrvpgsqrryrlatiansaigleprgdvdlgledescldgweyd 120  
 Qy 121 QDVLSTIVTWNLVCEDDWKAPLTISLFFVGLLGSFISQSLSDRFGKKNVLFVFMGMQ 180  
 Dd 121 kdiflsvtewnlvceedwktpltslffvglvllgsfsvsgqlsdrfgrkknvlfatmavq 180  
 Qy 181 TGFSLQIFSKNFEMFVFLVGVNGQISNYAAVFLGTEILGKSVRIIPSTLGVCIFYA 240  
 Dd 181 tgfslqifsknfemfvlvlgvngqisnyvafilgtelgksvriifstlgvctffa 240  
 Qy 241 FGYNVPLFAYFIRDRMLVALTMPGVLCAVLMWFIPESPRWLISQGRPEAEVIIRKA 300  
 Dd 241 lgymlplfayfirdwrmllaltvpcvlwfwfipesprwlisqrfreacdliqka 300  
 Qy 301 AKANGIVVPSTIFDPSELODSSKKQOSHNLDLRTWNIRMYTMSIMLWMTISVGYFG 360  
 Dd 301 akmsivapagifdplelqelnsikqkvildlfrtniativtmavmlwmltsvgyfa 360  
 Qy 361 LSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLLOVLPYRYSMATALFLCGSVLLFMQLV 420  
 Dd 361 lsldvplnhgdvynclslgllievpayftawllrtpryiaavlfwg99villfiqlv 420  
 Qy 421 PPDIYLIATVLMVGKFGVTAAFSMVYVYTAELYPTVVRNMGVSVSTASRLGSIISPYF 480  
 Dd 421 pedyfvsvigvlmgkfgitsafsmlyvftaelypvltrnmavgtstasrvgsiiapyf 480  
 Qy 481 VYLGAIDRFPLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 540  
 Dd 481 vylgaynrlmipyilmgsltlvltlftlfpesfgtvpdenleqmkgvkrfgck---kst 537  
 Qy 541 MLKDGQERPTILKSTAF 557  
 Dd 538 vsdreespkvl-itaif 553  
 RESULT 9  
 AAB20579  
 ID AAB20579 standard; Protein; 553 AA.  
 AC AAB20579;  
 XX  
 DT 11-DEC-2000 (first entry)  
 XX  
 DE Mouse OCTN1 amino acid sequence.  
 XX  
 KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;  
 KW identification; regulator; carnitine transport.

XX OS Mus musculus.  
XX PN WO200046368-A1.  
XX PD 10-AUG-2000.  
XX PF 04-FEB-2000; 2000WO-JP00619.  
XX PR 05-FEB-1999; 99JP-0028406.  
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX PA Nezu J, Ose A;  
XX PI WPI; 2000-586982/55.  
XX DR  
XX PT Organic cation transporter gene OCTN3 expressed in testis for  
PT identification of regulators of carnitine transport for use as drugs -  
XX PS  
XX Example 2; Fig 2; 58pp; Japanese.  
XX The present invention describes a mouse organic cation transporter  
CC protein (OCTN3). Also described are: (1) a method for screening  
CC compounds for their ability to regulate the transport of an organic  
CC cation into the cell, by generating a cell expressing OCTN3 at the  
CC cell membrane, contacting with the compound and organic cation, and  
CC observing the degree of transport of the organic cation; and (2) a  
CC method for screening compounds for their ability to be transported into  
CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell  
CC membrane, contacting with the compound and observing the degree of  
CC transport of the compound. OCTN3 can be used for the identification of  
CC regulators of the transport of organic cations (especially carnitine)  
CC into cells by OCTN3, for use as drugs. The present sequence represents  
CC the mouse OCTN1 amino acid sequence, which is used in an example from  
XX the present invention.  
XX SQ Sequence 553 AA;  
  
Query Match 74.5%; Score 2148; DB 21; Length 553;  
Best Local Similarity 72.2%; Pred. NO. 1.8e-225;  
Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;  
  
QY 1 MRDYEATLGEWGPQRILFFLLSIIIPNGFTGLSSVFLIATPEHRCRVPDAANLS 60  
DB 1 mrdyeviafgewgpqrilffllsaiipngfngmvsvflagtpehrcrlvdpdvnl 60  
QY 61 AWRNHTVPLRLRDGREVPHSCRRYRLATIANFSGALGLEPGRDVLGQLESCLDGWEFS 120  
DB 61 swrnhstplektdgrvppqscrryrlatiansamflepqgdvldleqescldgweyd 120  
QY 121 ODVYLSITVTENNLVCEDDNKAPITISLFFVGVLLGSFISQGLSDRGRKNVLFTVMGQ 180  
DB 121 kdifstivtenlvceddwktpittslffvlgcgsfvgglsdrfgrkvvfatmavq 180  
QY 181 TGFSEQLQFSNPFVFLVLCMGQISNVAAVFLGTTELCKSVRIIFSTGLGVCIFYA 240  
DB 181 tgfsvqlqfscnpevfvlvlgcmgqisnvvaavflgttelcksvrliifstglgvcffia 240  
QY 241 FGYVVLPLFAFYFIRDRLMLVALMPGVLCVAMWVFIPESPRWLISQGRFEAEVIRKA 300  
DB 241 igymvplfayfirdwrmllaltlplfcvplwlfipesprwlisrfaaeaiiqlka 300  
QY 301 AKANGIVVPSTIFDPSELDLSSKKQSHNLLDLRTWNTIMVIMLWVMSVGYFG 360  
DB 301 akmsivapagifdplelqelsikqkvildlfrtniatitvmavlmwmtsvgyfa 360  
QY 361 LSLDTPNLHGLDIFVNCFLSAMVEVPAYVLAWLLQLYLPFRYSMATALFLGSLVLLFMQLV 420  
DB 361 lslavpnlhgdvynclisgliepayftawllrtpryilagvlfwggvllliqvv 420  
QY 421 PPDLYYLATVLMVGKFGVTAFAFSNVVYVYTAELYPTVVRNMGVGSSTASRLGSILSPYF 480

DB 421 pedyfnfvsiglvmlgkfgitsafsmlyvftaelyptlvrmavgiitsmaarsvgilapf 480  
QY 481 VYLGAIDRFLPYILMGSILTILTAILTLFLPESFGTPLPDTIDQMLRVKGMKHKRTFSHTR 540  
DB 481 vylgaynrllipyilmgsltlglitiffbesfgvtlpenledmqkvrgfrcgk---kst 537  
QY 541 MLKDGQERPTILKSTAF 557  
DB 538 vsvdreespkvl-ita 553  
  
RESULT 10  
AAB43038  
ID AAB43038 standard; Protein; 560 AA.  
XX AC AAB43038;  
XX DT 08-FEB-2001 (first entry)  
XX DE Human ORFX ORF2802 polypeptide sequence SEQ ID NO:5604.  
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; erythematoidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX OS Homo sapiens.  
XX PN WO200058473-A2.  
XX PD 05-OCT-2000.  
XX PF 31-MAR-2000; 2000WO-US08621.  
XX PR 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
XX PI Shinkets RA, Leach M;  
XX WPI; 2000-602362/57.  
DB N-PSDB; AAC77247.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX Claim 11; Page 4791-4792; 5507pp; English.  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating

```
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergic, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
```

XX Sequence 560 AA;

SQ

Query Match  
Best Local Similarity 27.2%; Score 785.5; DB 21; Length 560;  
Matches 185; Conservative 105; Mismatches 212; Indels 47; Gaps 9;

Yy 17 FORLTFLLSIIIPNGFTGSSVFLIATPEHRCRP-----DAANLS 59  
|||:::||:||||: |||:: ||| |  
2 fqrlylcafnqslacghylasvmgvtphhvcrppgnvsqvfhnhnswnsedtgall 61  
|||:::||:||||: |||:: ||| |

Yy 60 SAWRNHTVPRLRDGR--EVPHSCRRYLATIANFSALGLEPRDVDLGQLBQSCLDGMW 117  
||:||||:||||: ||| | ||| | ||| | ||| | :||:  
Db 62 ssgkdyvvtqlngselwelwscrsnrkr----entsslgve---ytgskefcpvdgy 112  
||:||||:||||: ||| | ||| | ||| | ||| | ||| :

Yy 118 EFSQDYLVLTIVTEHNLCEDDKAPLTISLFVGVLGLSPFSGOLSRFGKKNVLFVTM 177  
||:||||:||||: ||| | ||| | ||| | ||| | ||| |  
Db 113 lyydntkwstavtqnvlvcdr:kwlamlqpifmgvgllsvtfgyfdrlgrvvlwats 172  
||:||||:||||: ||| | ||| | ||| | ||| | ||| |

Yy 178 GMQTGFSELPQRSKFEMPVVFLVLVGMGOISNYAAEVLGTIELKGVRITFSTLGVCI 237  
||:||||:||||: ||| | ||| | ||| | ||| | ||| |  
Db 173 ssmlflgiaafaydyttmaarfclamvasgylvgvymefigmksr-twasvhhs 231  
||:||||:||||: ||| | ||| | ||| | ||| | ||| |

Yy 238 FYAGYNVLPFAFYIRDW---RMLLVALTMPGVLCVALMWPIESPRLWLISOQRFEAE 294  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |  
Db 232 ffavgltlaigtvyrtwyqmilstvtpfilcc---wlpetpfwlisegryeeaaq 288  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |

Yy 295 VIIRAKANGIVPSTIFDPSELQDL-----SSKKQQSHNILLRTWNRMVTIM 346  
||:||||:||||: ||| | ||| | ||| | ||| | ||| |  
Db 289 kivdimakwn---rascklsellsldlqqpsnspetevkhnlsyfynwsitkrcit 344  
||:||||:||||: ||| | ||| | ||| | ||| | ||| |

Yy 347 SIMLMWTISGYFGSLDTPNLHGDI FNCFLSAMVEPAYVLAWLLLYLRFRYSMTA 406  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |  
Db 345 vwllwtfgsigfyfsinsvnlgneylnlfllggveipayfvfciamdkvgrrtvlays 404  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |

Yy 407 LFLGSVLLFMQLVPPDLYLIATYLVMVGKPGCVTAASFWMYYVAELYPTVVRRMGVCVS 466  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |  
Db 405 lfcsalacgvwmvfqkhqilgvvtamvgfaqaafglytytaelyptivrslavsg 464  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |

Yy 467 STASRGILSPYPVYLGAVDRFPYTLIMGSLTILTALLTLPLPSRTGPLPDIDQMRL 526  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |  
Db 465 smvrclasilapfsyddissiwifiqlfvgtumallsvgltkpetigklratweeaaq 524  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |

Yy 527 VKGMKHRRKT 535  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |  
Db 525 lesenesks 533  
||:||||:||||: ||| | ||| | ||| | ||| | ||| | ||| |

RESULT 11  
AAM00930  
ID AAM00930 standard; Protein: 584 AA.  
XX AC AAM00930;  
XX DT  
XX DE 01-OCT-2001 (first entry)  
XX DE Human bone marrow protein, SEQ ID NO: 406.  
XX DE Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;  
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;  
KW immunosuppressive; gene therapy; cytokine cell proliferation;

Db 241 gmsr-twasvhnshfavgvllvaltaqylvrtwlygmilstvtvpfilcc---wlppe 296  
QY 280 SPRWLISQGRFEAEVIRKAANKANGIVPSTIFDPSEIQLD-----SSKKQOSHNI 331  
Db 297 tpfwllsegryeaaqkivdimakwn---rasscklsellslldlqgvnsptevqkhl 352  
QY 332 LDLLRTWNRIMVTIMSLMWTISVGFGLSLDTNPNLHGDIFVNCFLSAMVEPAPVLAW 391  
Db 353 sylfynwsitkrtltwvlwftslgfsfslnsnlgnneylnlllgvveipaytfvc 412  
QY 392 LLQYLPRRYSMTALFLGGSVLLFQVLPPDLYLYLATVLMVGK-FGVTAAFSMVYVVT 450  
Db 413 iamdkvgrtvlayslfcslacqgvnmvpqkhyilgvvtamgkilpigaagilylt 472  
QY 451 AELYPTVVRNMGVSVSTASRLGSLSPYFVYLGAIDRFLPYILMGLSLTLTLLAILFLP 510  
Db 473 aelyptivrsilavsgsmvrcslapfsvdlssiwifipqifvgtmallsgvltklip 532  
QY 511 ESFGTPLPDTIDQMLRVKGMKHKRT 535  
Db 533 etlglkriattweeakiesenesks 557  
  
RESULT 12  
AAW64538  
ID AAW64538 standard; Protein; 554 AA.  
XX  
AC AAW64538;  
XX  
DT 21-OCT-1998 (first entry)  
XX  
DE Human liver cell clone HP01293 protein.  
XX  
KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;  
KW differentiation; immune system; stimulator; suppressor; regulator;  
KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;  
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO9821328-A2.  
XX  
PD 22-MAY-1998.  
XX  
PF 07-NOV-1997; 97WO-JP04056.  
XX  
PR 13-NOV-1996; 96JP-0301429.  
XX  
PA (PROT-) PROTEGENE INC.  
XX  
PA (SAGA ) SAGAMI CHEM RES CENTRE.  
XX  
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
XX  
XX WPI; 1998-297932/26.  
DR N-PSDB; AAV49558, AAV49559.  
XX  
XX Human protein having transmembrane domain - useful for, e.g.  
PT research and nutrition  
XX  
PS Claim 1; Page 96-98; 205pp; English.  
XX  
CC AAW64534-W64558 represent human proteins containing a transmembrane  
CC domain. These proteins can be used for, e.g. research and nutrition, and  
CC may have cytokine and cell proliferation/differentiation, immune  
CC stimulating/suppressing, haematopoiesis regulating, tissue growth,  
CC activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,  
CC receptor/ligand, anti-inflammatory or tumour inhibition activity.  
XX  
SQ Sequence 554 AA;

Query Match 25.5%; Score 735; DB 19; Length 554;

Best Local Similarity 35.8%; Pred. No. 3.8e-71;  
Matches 196; Conservative 86; Mismatches 216; Indels 50; Gaps 12;  
QY 1 MRDYDEVTAFLGEMGPFOR---LTFLLSASLIIPNGFTGLSSVFLIATPEHRCRVPPDAAN 57  
Db 1 mptvddilleqygeswfgkqafllilclisaafap---icvgivflgtphdhcspgvae 57  
QY 58 LSS--AWR-----NHTVPLRLRDGREVPHSCRRYR-----LATIA-NFSAIG 96  
Db 58 lsqrcgwspaeelnlyvvp9lpageaflgqcrreyevdwngalscvdpplasiatnrshp 117  
QY 97 LBPGRDVLGQLEQBSCLDGMWFEFSQDYYLSTIVTEWNLVCEDDMKAPLTISLFFVGVLLG 156  
Db 118 lqp-----cqdgwvy--dtpgssivtefnlvcadswklidfqsclnagffg 162  
QY 157 STISQSLSDRFGKNNLVFTVMGMQTFSELFQIFSKNFEMFVVLVVGMGQISNVYAAV 216  
Db 163 sigvgyfadrfrgrklcligtvlvnavsgvlmafspnysmlfrllqglvskgnmagyt 222  
QY 217 LGTEILGKSVRIESTLGLGVIFYAF--GYMVLPLFAYFIROWRMLLVALTWPGVLCVALM 274  
Db 223 litefvsgsr---rtvaimygmavtvgvalgtlayalphwrwlqlavslptflilly 279  
QY 275 WFIPESPRWLISQGRFEAEVIRKAANKANGIVPSTIFDPSEIQLDSSKKQOSHNI 334  
Db 280 wcvpesprwllsqkrnteaikimdhiaqngklppadikmlsleedvtek--lspsfadl 337  
QY 335 LRTWNRIMVTIMSLMWTISVGFGLSLDTNPNLHGDIFVNCFLSAMVEPAPVLAWLL 394  
Db 338 fitprlrkrftlmylftdsvlyqgllhmgatsgnlyldflysalveipgafialiti 397  
QY 395 QYLPTRYSMATALFLGGSVLLFQVLPPDLYLYLATVLMVGKFGVTAAFSMVYVTAELY 454  
Db 398 drvgriypmavnllagaacvlmifispldhwnliimcvgmrgitaiqmclvnaely 457  
QY 455 PTVVRNMGVSVSTASRLGSLSPYFVY-LGAYDRFLPYILMGLSLTLTLLAILFLPESF 513  
Db 458 ptfvrnlngvmvcsslcdiggitpfivfrlvewqalpilfavgliaagvtililpetk 517  
QY 514 GTPLPDTI 521  
Db 518 gvalpetm 525  
  
RESULT 13  
ARB49401  
ID AAB49401 standard; Protein; 540 AA.  
XX  
AC AAB49401;  
XX  
DT 02-MAR-2001 (first entry)  
XX  
DE Murine organic anion transporter 6.  
XX  
KW Murine; organic anion transporter 6; mOAPP6; cancer; inflammation;  
KW cardiovascular disease; central nervous system disorder; kidney disease;  
KW liver disease; autoimmune disease.  
XX  
OS Mus sp.  
XX  
PN WO200070048-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 15-MAY-2000; 2000WO-US13316.  
XX  
PR 14-MAY-1999; 99US-0134137.  
PR 12-MAY-2000; 2000US-0570293.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Feild J, Yue L, Ellens H;









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 10:10:16 ; Search time 23.33 Seconds  
(without alignments)  
537.262 Million cell updates/sec

Title: US-09-521-195-3  
Perfect score: 2883  
Sequence: 1 MRDYDEVTAFLGNGPQRL.....HTRMLKDGQERPTILKSTAF 557

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	726	25.2	555	3	US-08-501-572-3
2	726	25.2	555	3	US-09-040-444-3
3	718	24.9	556	3	US-08-501-572-1
4	718	24.9	556	3	US-09-040-444-1
5	714.5	24.8	553	3	US-08-501-572-2
6	714.5	24.8	553	3	US-09-040-444-2
7	652.5	22.6	537	2	US-08-647-397-2
8	302.5	10.5	520	4	US-08-964-127-2
9	234	8.1	494	2	US-09-031-392-5
10	234	8.1	494	4	US-09-299-549-5
11	214.5	7.4	286	4	US-08-964-127-4
12	208	7.2	492	2	US-08-355-844-3
13	208	7.2	492	5	PCT-US95-16126-3
14	203	7.0	493	2	US-09-031-392-10
15	203	7.0	493	4	US-09-299-549-10
16	198	6.9	109	2	US-08-647-397-4
17	193	6.7	509	2	US-09-031-392-6
18	193	6.7	509	4	US-09-299-549-6
19	186	6.5	500	2	US-09-031-392-7
20	186	6.5	500	4	US-09-299-549-7
21	184	6.4	563	4	US-09-031-392-2
22	184	6.4	563	4	US-09-299-549-2
23	175	6.1	488	2	US-08-928-692-11
24	171	5.9	322	4	US-08-964-127-6
25	168.5	5.8	534	2	US-09-031-392-4
26	168.5	5.8	534	4	US-09-299-549-4
27	155	5.4	383	2	US-09-031-392-3

28	155	5.4	383	4	US-09-299-549-3	Sequence 3, Appli
29	153.5	5.3	524	2	US-08-928-692-12	Sequence 12, Appl
30	149	5.2	584	2	US-08-928-692-13	Sequence 13, Appl
31	131	4.5	488	2	US-08-928-692-10	Sequence 10, Appl
32	110	3.8	473	1	US-08-597-236-13	Sequence 13, Appl
33	110	3.8	473	1	US-08-746-682A-13	Sequence 13, Appl
34	104.5	3.6	1835	3	US-08-836-325-15	Sequence 15, Appl
35	101.5	3.5	341	1	US-08-423-564-5	Sequence 5, Appli
36	97	3.4	342	3	US-08-785-928-1	Sequence 1, Appli
37	97	3.4	342	3	US-08-728-603-17	Sequence 17, Appl
38	97	3.4	916	2	US-08-928-692-58	Sequence 58, Appl
39	94.5	3.3	390	5	PCT-US92-02091-8	Sequence 8, Appli
40	93.5	3.2	467	2	US-08-805-118-3	Sequence 3, Appli
41	93.5	3.2	480	2	US-08-724-394A-9	Sequence 9, Appli
42	93	3.2	560	2	US-08-805-118-4	Sequence 4, Appli
43	92.5	3.2	419	1	US-08-385-186-2	Sequence 2, Appli
44	92.5	3.2	568	4	US-08-637-823B-27	Sequence 27, Appl
45	92.5	3.2	584	4	US-08-637-823B-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1  
US-08-501-572-3  
; Sequence 3, Application US/08501572  
; Patent No. 6063623  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; TITLE OF INVENTION: Transport protein Which Effects The  
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,  
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson,Parabow,Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/501,572  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toohy, Kimberlin M  
; REGISTRATION NUMBER: 35,391  
; REFERENCE/DOCKET NUMBER: 02481.1453-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 555 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-501-572-3

Query Match 25.2%; Score 726; DB 3; Length 555;  
Best Local Similarity 35.0%; Pred. No. 5.6e-68;  
Matches 187; Conservative 98; Mismatches 223; Indels 26; Gaps 14;

```

1:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    6 DDVLEHGGEFFHFKQMFLLALLSATFAP1--YVG1--VFGLFTPDHRCRSPGVAELSLR   62
Qy    60 SAWR-----NHTVFLRLRDGVEPHSCRRYLAA-TIANFSALGLEPCGRDVDLGOLEQESC 113
1:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    63 CGWSPAEBLNTYVFGPGPAGEASPCRERRYEVDNQNSTFDCCVPDLASLDINRSRLPLGPC 122
Qy    114 LDGHEFSQDVYLSITIVTEWNLCEDDKKAPTISLFVFGVLLGSFISQLSDRFRGNVL 173
11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    123 RDGWYV---ETPGSSIVTFENILVCANSWMULDLFQSSVNUGFFIGSMSTGYIADRGRKLCL 180
Qy    174 FVTMGMTGFSEFLQIFSKNFEMFVVFLVLMGMGIISNVAAAFVLGTILCKSVRIIFSTL 233
11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    181 LTTVLINAAGVLAISPTTYTWMLIFRLIOGLYSKAGWLIGYLITEFVGGRYR---RTV 237
Qy    234 GVC--IFYAFGYMYDLPFAFYFIROWRMLLVALTMPGVLCVALMWFIPESPRWLISOGRFE 291
11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    238 GIFQVAYTVGLLVLAGVAYALPHWRLOQTVALPNPFLLYYWCYIPESPRWLISOQNKA 297
Qy    292 EAEVIIRKAAKANGIVPSTFDPSQLDLSKKQOOSHNLDLLRTWNIRMVIMSIMLW 351
11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    298 EAMRLIKHIAKKNCKSLPASL-QBLRLEEETGKK-LNPSELDLVRTPQIRKHMTMIWYNW 355
Qy    352 MTISVGVEGLSDTPNNLHGD-IFVNCFLSAMVEYPAYVLAWLLLOVLPTRYSMATAFLG 410
11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    356 FTSSVLVQGLTMHN-GLAGDNIVLDFEYSALVEFPAAFMIIIDRIGRPWAASNMA 414
Qy    411 GSVLLEMOLPPDLYLATVLVMVGFGVTAARSMVYVYTAEYLPVVRNMVGCVSSTAS 470
11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    415 GAACLASVFPDGLDWLKLIISCLGRGGITWATEIVCLVNAAELYPIFRLNGVHCSSMC 474
Qy    471 RLGISLFPYFYV-LGAYDRFLPYLMGSLTILTALTFLFLPESFGTPLDTIQD 523
11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1      11:1
Db    475 DIGGIITPFLVYRLTNWLEPLMWFGVLGVLGAGGLALLPETKGALPETIEE 528

RESULT      2
US-09-040-444-3
; Sequence 3, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use
```

```

; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-040-444-3

```

Query Match	25.2%;	Score 726;	DB 3;	Length 555;
Best Local Similarity	35.0%;	Pred. No. 5.6e-68;		
Matches 187;	Conservative	98;	Mismatches 223;	Indels

Qy	5	DEVTAFLGKGPFORLITF---LLSASIIPNGFTGSSVFLIATPHECRCPVDAANLS--59
Db	6	DDVLEHGGEFFHFKQOKMFFLLALISATFAP-I-XVGI--VPLGTFPHDRKSPGVAELSLR 62
Qy	60	SAWR-----NHTVPLRLRDGVEPHSCRRYRLA-TIANFSALGLEPGRVDLGLQEQESC 113
Db	63	CGWSPAEBELNYTPGCPAGEASPRQCRRRYEVDWNOSTFCDVDPLASLDTNRSRLPLGPC 122
Qy	114	LDGWEFSQDVIYLSITTEWNLCVEDDMKAPLTTLSLFVGVLLGSISGQLSDRGRKNVL 173
Db	123	RDGVVY--ETPGSISVTEFNLVCANSMLDLFOSSNVGVFFIGMSIGYIADRGRKLC 180
Qy	174	FVTMGMTGFSFIQISKNFEMFWLFLVLMGMSINYAAAFVLGTEILGKSVRIIFSTL 233
Db	181	LTVTVLINAAAGVLMASPTVTWMLIFPLIOGLVSKAGWLIGYLIITFVGGRYR---RTV 237
Qy	234	GVC---IYAFGYVMVLPFAFYIRDRMLLVALMPGVLCVALMWFIPESPRMLISQGRE 291
Db	238	GIFYQVAYTVGLLVLAGVAYLAPHWRWLOFTVLPNFFFLLYWCYIPESPRMLISONKNA 297
Qy	292	EAEVIRKAAKANGIVPSTIFDPSELDQLSSKKQOSHNTLDLRTWNRTWMTINSLMW 351
Db	298	EAMRIIKHIAKKNKSLPASL-QRLREETGKK-LNPSELDVLRPTQKKTWILWYMW 355
Qy	352	MTISVGYFGLSLDTPNLHGD-IPVNCFLSAMVEVPAYVLAWLLQYLPRRYSMATALFTLG 410
Db	356	FTSSVLYQGLIMHM-GLAGDNIIYLDFFYSALVEFPAAFMTIILDRGRYPWAAASNMYA 414
Qy	411	GSVLLPMQLVPPDLYLATVLVWVGFGFYTAFAFSMVYVYTAELYPTVNVMGVGSSTAS 470
Db	415	GAACLASVFPDGLQWLKIIISCLGRWGITMAYEIVCLVNAELYPFFIRNLGVHICSSMC 474
Qy	471	RLGSILSPFVY-LGAYDRFLPYLMGSIITLAILTLFLPESFGTPLDPTIDQ 523
Db	475	DIGGIITPFLVRLTNLWLEPLMWPGVLGVLVAGGLALLPRTKGALPETTEE 528

RESULT 3  
US-08-501-572-1  
; Sequence 1, Application US/08501572  
; Patent No. 6063623

GENERAL INFORMATION:  
 APPLICANT: Koepsell, Hermann  
 APPLICANT: Grundeman, Dirk  
 APPLICANT: Gorboulev, Valentin  
 TITLE OF INVENTION: Transport protein which Effects The  
 TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,  
 TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
 STREET: 1300 I Street, N.W., Suite 700  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3315  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0. Version #1.30

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohay, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4400
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-501-572-1

```

```

Query Match 24.9%; Score 718; DB 3; Length 556;
Best Local Similarity 34.0%; Pred. No. 3,9e-67;
Matches 194; Conservative 96; Mismatches 229; Indels 52; Gaps 14;

```

```

QY 1 MRDYDEVTAFLGEGWGPQRILFFLL---SASIIPIGFTGLSSVFLIATPEHRCRVDPDAAN 57
DB 1 MPTVDDVLEQGVGEFGWFOQAFLILCLISASLAPI-YVGI--VFLGFTPGHYCQNGVAE 57
QY 58 LSS--AWR-----NHTVP-LRLRDGREVPHSCRRYL-----ATIANFSAL 95
DB 58 LSQRCGWSQAEELNYTVPLGSPDEASFLSQCMRYEVDWQSTLDCVDPSSLVANRSOL 117
QY 96 GLEPGRDVLDLQLEQESCLDGEFSDQVYLSITVTEWNLVCEDDWKAPLTIISLFFVGVLL 155
DB 118 PLGP-----CEHGWY--DTPGSSIVTEFNLCVGDANKVDLFOSCVNLGFFL 162
QY 156 GSFTSGQLSDRFRGKKNVLFVTMGMTGFSFLOIFSKNFEMFVFLVVGNGOISNYVAAF 215
DB 163 GSVLVGYIADRFGRKLCLLVTLTVTSVSGVLTAAPDYTSMLLFRLLQGVMSKGSWVSGY 222
QY 216 VLGTIELGKSVRIIFSTLGVCIYAF--GYMVLPLFAYFIRDMRMLLVALTMPGVLCVAL 273
DB 223 TLITFEVGSYR---RTAILYQMAFTVGLVGLAGVAYAIPTDRLQLAVSLPTFLFLY 279
QY 274 WVFIPESRWLLISQGRFEEAEVIRKAANGIIVVPSTIFDPSLQDLSKKQOOSHNLID 333
DB 280 YWFPESRWLLISQGRFEEAEVIRKAANGIIVVPSTIFDPSLQDLSKKQOOSHNLID 337
QY 334 LLRTWNIRMTVIMSLMWTISVGYFGLSLDTPNLHGDIIVFNCFLSAMVEVPAYVLAWL 393
DB 338 LFRTPNLRKHTVILMYLWFSQAVLYQGLIMHVAGATGANLYLDFFYSLSVEFPAAFIILVT 397
QY 394 LQYLPRIYSMATLFLGGSVLLFMQLVPPDLYLYLATVLMVVGKFGVTAAFSMVYVYTAEL 453
DB 398 IDRIGRIYPIAASNLVGTGAACLLMIFIPHEHLNLTACLRMGATVILEMVCVNAEL 457
QY 454 YPTVVRNMGVSGVSTASRLGSLSPFYV--LCAVDRLPYLIMGSLILITLTLFIPES 512
DB 458 YPTFTRNLGMVCSALCDLGIPTFMVFRMLMEVQALPLILFGLVLTAGAMTLLLPET 517
QY 513 FCTPLPTIDQMLRVKGMKHKRTPSHTRLK 543
DB 518 KGVALPTEIEAENL-GRKSKAKENTYILQ 547

```

```

RESULT 4
US-09-040-444-1
; Sequence 1, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk

```

```

; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-040-444-1

```

```

Query Match 24.9%; Score 718; DB 3; Length 556;
Best Local Similarity 34.0%; Pred. No. 3,9e-67;
Matches 194; Conservative 96; Mismatches 229; Indels 52; Gaps 14;

```

```

QY 1 MRDYDEVTAFLGEGWGPQRILFFLL---SASIIPIGFTGLSSVFLIATPEHRCRVDPDAAN 57
DB 1 MPTVDDVLEQGVGEFGWFOQAFLILCLISASLAPI-YVGI--VFLGFTPGHYCQNGVAE 57
QY 58 LSS--AWR-----NHTVP-LRLRDGREVPHSCRRYL-----ATIANFSAL 95
DB 58 LSQRCGWSQAEELNYTVPLGSPDEASFLSQCMRYEVDWQSTLDCVDPSSLVANRSOL 117
QY 96 GLEPGRDVLDLQLEQESCLDGEFSDQVYLSITVTEWNLVCEDDWKAPLTIISLFFVGVLL 155
DB 118 PLGP-----CEHGWY--DTPGSSIVTEFNLCVGDANKVDLFOSCVNLGFFL 162
QY 156 GSFTSGQLSDRFRGKKNVLFVTMGMTGFSFLOIFSKNFEMFVFLVVGNGOISNYVAAF 215
DB 163 GSVLVGYIADRFGRKLCLLVTLTVTSVSGVLTAAPDYTSMLLFRLLQGVMSKGSWVSGY 222
QY 216 VLGTIELGKSVRIIFSTLGVCIYAF--GYMVLPLFAYFIRDMRMLLVALTMPGVLCVAL 273
DB 223 TLITFEVGSYR---RTAILYQMAFTVGLVGLAGVAYAIPTDRLQLAVSLPTFLFLY 279
QY 274 WVFIPESRWLLISQGRFEEAEVIRKAANGIIVVPSTIFDPSLQDLSKKQOOSHNLID 333
DB 280 YWFPESRWLLISQGRFEEAEVIRKAANGIIVVPSTIFDPSLQDLSKKQOOSHNLID 337
QY 334 LLRTWNIRMTVIMSLMWTISVGYFGLSLDTPNLHGDIIVFNCFLSAMVEVPAYVLAWL 393
DB 338 LFRTPNLRKHTVILMYLWFSQAVLYQGLIMHVAGATGANLYLDFFYSLSVEFPAAFIILVT 397
QY 394 LQYLPRIYSMATLFLGGSVLLFMQLVPPDLYLYLATVLMVVGKFGVTAAFSMVYVYTAEL 453
DB 398 IDRIGRIYPIAASNLVGTGAACLLMIFIPHEHLNLTACLRMGATVILEMVCVNAEL 457

```



Best Local Similarity 35.4%; Pred. No. 9.1e-67;  
Matches 194; Conservative 87; Mismatches 216; Indels 51; Gaps 13;

```

QY 1 MRDYDEVTAFLGECGPFOR---LIFFLLSASLIIPNGFTGLSSVFLIATPEHRCRVPDPAAN 57
Db 1 MPTVDILQEGWESGFWKQAFILCLLSAFAF---ICVGIVFGLTFTPDHHCOSPGVAE 57
QY 58 LSS---AWR-----NHTVPLRLRDGREVPHSCRRYR-----LATIA-NFSALG 96
Db 58 LSQRCGSPAEELNVTVPCLGAGEAFGLQCRRYEVDMNQSALSCVDPPLASLATNRSLP 117
QY 97 LEPGRVDVLGOEBCSDUDNEFSDVYLSITVTENNIVCEDDKAPLTISLFFVGVLLG 156
Db 118 LGP-----CQDQWVY--DTPGSSIVTEFNLCVADSNKLDLFQCLNAGFFFG 162
QY 157 SPISGOLSDRGRKNVLFVTMGOTGFSFLOIFESKNEMFVFLVYVGMGOISNYAAV 216
Db 163 SLUGYGFADRFGRKLCGLGTVLNNAVSVLMFSPNYMSMLLFLLQGLVSKGNWAGYT 222
QY 217 LGTEILKSVRIIFSTLGVCIYAF--GYMVLPLFAFIRDMRLVLTMPGVLCVALW 274
Db 223 LITEFVSGSR---RTVAIMQMAFTVGLVALTGLAYALPHRWLQLAVSLPTFLFLY 279
QY 275 WEIPSPRWLISQGRPERAEVIRKAANGIVVPSTTFDSELDSSKKQKQSHNILD 334
Db 280 MCVPESPRWLLSQKRNTEAIKIMDHIAQKNGKLPADLKMLSLEEDVTEK--LSPSEADL 337
QY 335 LRTNIRMTVIMSIMLWNTISVGYEGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLL 394
Db 338 FTPLRLKRTFTILMYLWFTSDVLYOGLLHMGATSGNLYLDFLSALVEIPGAFIALITI 397
QY 395 QYLPYRYSMATLFLGGSVLLFMQLVPPDLYLATVLMVVGKFGVTAAFSMVYVYTAELY 454
Db 398 DRVGRIYPMAYNSLLAGAACLVIFISPDHLNIIIMCVGEMGITTAQIMCLVNAELY 456
QY 455 PTVWRNMGVGSSTASRLGSLSPYFY- LGAYDRFLPYILMGSJLTILTALTILFLPESF 513
Db 457 PTFVRLNRMVWGSSICDGIITPFIPLRVLMWALPLILFAVLGLAAGVTLILPETK 516
QY 514 GTPLPDTI 521
Db 517 GDALPETM 524

```

RESULT 7  
US-08-647-397-2

```

; Sequence 2, Application US/08647397
; Patent No. 5972702
; GENERAL INFORMATION:
; APPLICANT: Beier, David R.
; APPLICANT: Brady, Kevin P.
; TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,397
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7048

```

TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-647-397-2

```

Query Match 22.6%; Score 652.5; DB 2; Length 537;  
Best Local Similarity 29.9%; Pred. No. 3e-60;  
Matches 167; Conservative 104; Mismatches 246; Indels 41; Gaps 12;

```

QY 4 YDEVTAFLGEGWGFQRLIFLLSASLIIPNGFTGLSSVFLIATPEHRCRVPDPAANLSSAWR 63
Db 3 FSEILDRVSGMGPFQYLHVTLALPILGIANHNLLOIFTATTDDHRCRPPNASL-EPW- 60
QY 64 NHTVPLRLRDGREVPHSCRRY---RLATIANFSALGLEPGRDVLQGLEQESCLDGWEFS 120
Db 61 --VLPLG-PNGK--PEKCLRFVHLPHNASLPNDTQGATEP-----CLDGWIYN 102
QY 121 QDYVLSTIVTENNIVCEDDKAPLTISLFFVGVLLGFSISGQLSDRFRGNVLFVTMGMQ 180
Db 103 ST--RDTIVTWDLVCGSNKLEMAQSVFMAGILVGPVFGELSDRGRKPILTWSYLLL 160
QY 181 TGRSFLOIFSKNEMFVFLVYVGMGOISNYAAVFLGTLEILKSVRIIFS-TLGVCIYF 239
Db 161 AASGSAFAFSPSLTVYIMIFRFLCGCSISGISLSTIILNVEWVPTSTRAISSTTIGYC--Y 218
QY 240 AFGYMYLPLFAFIRDMRLVLTMPGVLCVALWFIPESPRWLISQGRPERAEVIRK 299
Db 219 TICQFPLGLAYAVPQWRWQLSVSAAFIFLSLWNVPESIRWLVLSKFSRALKTLQR 278
QY 300 AAKANGIVPSTTFDPSL-----QDLSSKKQKQSHNILDLLRTNIRMTVIMSIMLWMTI 354
Db 279 VATFNGKKKEGKLTVEELKFNLDKITSQVAKV-YGLSDLFVSLIRRVTFCLSLAWFAT 337
QY 355 SVGYFGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLLOYLPRRYSMATLFLGGSVL 414
Db 338 GFAYYSLANGVVEEFGVNIYILQIFGGVDIPAKFETILSISYLGRRITQGLLILAGVAI 397
QY 415 LFMQLVPPDLYLATVLMVVGKFGVTAAFSMVYVYTAELYPTVVRNMGVGSSTASRLGS 474
Db 398 LALIFVSSEMQLLRTALAVEGKGLSGFSCLFLYTSELYPTVLRGTGMSINIRAVGS 457
QY 475 ILSPYFVYLGAIDRFPLPYILMGSJLTILTALTILFLPESFGTLPDITID-----QMLRV 527
Db 458 MIAPLVKITGELQFPFIPNVIFWTMLIGSSAAFFLETLNRPLPETIEDIQDWYQQTKKT 517
QY 528 KGMKHKRTPSHTRMLKDG 545
Db 518 KQPEAEKASQTIPLKTG 535

```

RESULT 8

```

US-08-964-127-2
; Sequence 2, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/964.127  
FILING DATE: 06-NOV-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crews, Ph.D., L. Lee  
REGISTRATION NUMBER: P-43,567  
REFERENCE/DOCKET NUMBER: 07334/038001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 520 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-964-127-2

Query Match 10.5%; Score 302.5; DB 4; Length 520;  
Best Local Similarity 22.5%; Pred. No. 2.5e-23;  
Matches 125; Conservative 98; Mismatches 231; Indels 101; Gaps 20;

QY 40 VFLIATPEHRCVRPDAANLSSAARNHTVPLRLDROGVEPHSCRRYRLATIANFSAALGLEP 99  
Db 6 ITFLAPLH-CHYGAPPNAGWEPNAGSVASAAALASASRVATSDPSCSGFAP 64  
QY 100 GRVDLGLQEQSCDGEFSDVYLSI-IVTEWNLVGEDDKWAPLITISLFFVGLGSF 158  
Db 65 P-----DFNHCLKWDYNGPLVTTNAIGQWDLVCDLQWVILEQILFILGFASSYL 116  
QY 159 ISGQLSDRFGKRVLFVTMG-----QTGSFLOIFSKNPFEMVFLVVLGM 205  
Db 117 FLUGPADRFGRGVLLTGLVGPVCGVGAAGSTGYMALR-----FLLGFLLAGV 168  
QY 206 GOISNKVAAFVLGTILCKSVRIIFSTLGVCIYAFVAGYMWLPLFAYFTDRWRMLLVALTM 265  
Db 169 D-----LGWLMRLPLCDPTQRLRVALAGELVGGHFLGL-ALYSKQWRFLORMITA 222  
QY 266 PGVLCVALW--FIPESPRWISQGRFEAEVIRKAAKAN---GIVVPSTIFDPSELOD 320  
Db 223 PCILFLFYGWPGLFESARMLIVKRIEQAQSVLRILAEARNRPHQMLGEEA--QEAALQD 280  
QY 321 LSSKKQOSHNLDRTNIRWVTIMTM-LWMTISVGYFGLSLDTPNLHGDIFVNC--- 376  
Db 281 L-----ENCPCPATSFSAFLNTRNWKNL-----LILGTFNTIAHAIHCYQP 327  
QY 377 -----FLSAMVEVPVAVLWMLLQYLPVRYSMATLFLGSGVLLFMQLVPPDLY- 425  
Db 328 VGGGSGSPDFVLCSSLAGSTAALACVFLGVTVDFRGRIILLNMTLGTIASLVLLGLWD 387  
QY 426 YL-----ATVLVWVGFGVTAFAFVSVYVYTAELPTVVRNMGVSSVTSRIGSLSP-YF 480  
Db 388 YLNEAAITFVGLGFSQAAAILSTLAAAEVPTTVRGRGLGIMALGALGSLGSPAQR 447  
QY 481 VYLGYDRFLPYLMLGSLTILTAITLFLPESFGTLPDITDQMLRVKMKHRRTPSHR 540  
Db 448 LHMG-HGAFLOHVLAACALICLSIMLLPETKRLDPE----- 485  
QY 541 MLKDGQ--ERPTLK 553  
Db 486 VLVDGELCRPSLLR 500

RESULT 9  
US-09-031-392-5  
Sequence 5, Application US/09031392  
Patent No. 5942398  
GENERAL INFORMATION:  
APPLICANT: Tartaglia, Louis A.  
APPLICANT: Weng, Xun  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES  
ENCODING GLUTEX AND USES THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/031,392  
FILING DATE: 26-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Meiklejohn, Ph.D., Anita L.  
REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 07334/072001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-031-392-5

Query Match 8.1%; Score 234; DB 2; Length 494;  
Best Local Similarity 24.7%; Pred. No. 3.9e-16;  
Matches 112; Conservative 85; Mismatches 158; Indels 98; Gaps 23;

QY 144 LTISLFFVGLGSFISGQLSDRFRK-----NVLFVTMGMTGTGFSFLOIFSKNPFEMV 197  
Db 65 LSAIFSVGMIGSFVGLFVNFRGRNMLIVNLLAIAGGCLMGFCI--AESVEMLI 121  
QY 198 VLFVLGM--GOISNKVAAFVLGTILCKSVRIIFSTL-----GVCIFYAGYMWLPL 248  
Db 122 LGLRIIGLFCGLCTGFPVYI--GEISPTALRGAFGLTQLGIVIGLVAQIFGLKVI-- 177  
QY 249 FAYFIRD-WRMILLVALTMFGLCVLWLFIPESPRW-LISQGRFEAEVIRKAAKANGI 306  
Db 178 --LGTEDLWPLLITPAIQCAALPFCPSPRELLINRKEEKAKELQRLWGTEDV 235  
QY 307 VVSTFDSELODLSKKQOSH--NILDLRTWNRWVTIMSIMLWMT-----ISVGY 358  
Db 236 AQ-----DIOEMKDESMRMSQEQVTVLELFRAPNYRQPIIISIMLIQSOQLSGINAVY 290  
QY 359 FGLSLDTPNLHGDIFVNCFLSAMVEVPAY-----VLAWLLQYLPVRYSMAT 405  
Db 291 YSTGI-----FKDAGVQEPVATIGAVVNTIFTVSVFLVERAGR--TLH 335  
QY 406 ALFLQG-----SVLLFMQLVPPDLY-----YLATVLVWVGKFGVTAFAFVSVYVYTAELYP 455  
Db 336 LIGLGMAFCSILMTISLLKONYSWMSFICAILLVFAFFEIGPG-PIPWFIIVALEFG 394  
QY 456 TVVRNMGVSS---TASRLGSLSPYFV-YLGYADRFPLPYLMLGSLTILTAITLFL-L 509



Db 395 QGPRPAAVAGCSNWTNLFVGLLFPSPATFYLGA-----VFIVTFVLFVIFVFTFFKV 450  
QY 510 PSECTPLPDIDQMLRV-----KGMKHKRTP 536  
Db 451 PETRGR---TFEITRAFEGVQTGTGRGKGP 479

RESULT 10  
US-09-299-549-5  
; Sequence 5, Application US/09299549  
; Patent No. 6136347  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/299,549  
; FILING DATE: 26-APR-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/031,392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 494 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-299-549-5

Query Match 8.1%; Score 234; DB 4; Length 494;  
Best Local Similarity 24.7%; Pred. No. 3.9e-16;  
Matches 112; Conservative 85; Mismatches 158; Indels 98; Gaps 23;

QY 144 LTISLFFVGVLLGSPISQSLDRGK-----NVLFVTMGMTQGFSLQIFSKNFEMFV 197  
Db 65 LSAVIFSVGGMIGSSVGLFNRFRGRNSMLVNLATAIAGGLMGFKCI---AESVEMLI 121

QY 198 VLFVLVGM--GQISNVAAVFLVGTILCKSVRIIFSTL-----GVCIFVAFGYMVLPL 248  
Db 122 LGRLLIGFCGLCTGFVPMYI--GEISTALRGAFGLNLQIGVIGILVAQIFGLKVI-- 177

QY 249 FAYFIRD--WRMLNVALTPGVLCAVWFWIPESRW-LISQGRFEAEVIRKAAKANGI 306  
Db 178 --LGTEDLWPLLGLTILPAIIQCAALPCFSPFLLINKKEEKAKELIQLRWGTEDV 235

QY 307 VVPSTIFDPSELDUSSKKQSH--NILDLLRTWNIRMTMSIMLWMT-----ISVGY 358  
Db 236 AQ-----DIQEMKDESRMSQEKQVTVLELFRAPNYROPILIIISIMQLSQSLGAINAVFY 290

QY 359 FGLSLDTNLHCDIFVNCFLSAMVEVPAY-----VLAWILLQLVLPKRYSMAT 405

Db 291 YSNGI-----EKDAGVQEPVATIGAGVVNFIETVWSVFLVERAGRR--TLH 335  
QY 406 ALPLGG-----SVLLFMQLVPPDLY-----YLATVLVWVGKEGVTAAAFSMVTVYTAELYP 455  
Db 336 LIGLGMAFCISLMTISLLKONYSWMSFCIGAILLVFAFFEIGPG-PIPWFIVAELFG 394  
QY 456 TVVRNMGVGVSS---TASRLGSILSPYFV-YLGAYDRELPLPYILMGSILTILTAITLFL-L 509  
Db 395 QGPRPAAVAGCSNWTNLFVGLLFPSPATFYLGA-----VFIVTFVLFVIFVFTFFKV 450  
QY 510 PSECTPLPDIDQMLRV-----KGMKHKRTP 536  
Db 451 PETRGR---TFEITRAFEGVQTGTGRGKGP 479

RESULT 11  
US-08-964-127-4  
; Sequence 4, Application US/08964127  
; Patent No. 6277565  
; GENERAL INFORMATION:  
; APPLICANT: Grangearl, Andrew David John  
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE  
; TITLE OF INVENTION: MOLECULES  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/964,127  
; FILING DATE: 08-NOV-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crews, Ph.D., L. Lee  
; REGISTRATION NUMBER: P-43,567  
; REFERENCE/DOCKET NUMBER: 07334/038001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 286 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-964-127-4

Query Match 7.4%; Score 214.5; DB 4; Length 286;  
Best Local Similarity 24.6%; Pred. No. 1.9e-14;  
Matches 69; Conservative 51; Mismatches 122; Indels 39; Gaps 8;

QY 40 VFLIATPEHRCRYPDPAANLSSAMRNHTVPLRLRDGREVPHSCRRYRLATIANFSAIGLEP 99  
Db 6 IFTIATPLH--CHYGATPPNAGSGWGPNNAGSVGSAALAAASRAVATSTDFSCSGFAP 64

QY 100 GRDVDLQGLQESCLDQWFSQDVYLST-IVTEWNLVCEDDMKAPLITISLFFVGVLLGSF 158  
Db 65 P-----DFNCLDKWDYNGLPLVLTNAIGQWDLVCDLGNQVILEILFILGFASGYL 116



```

; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031.392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meikiejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07304/072001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-031-392-10

Query Match 7.0%; Score 203; DB 2; Length 493;
Best Local Similarity 25.0%; Pred. No. 7.3e-13;
Matches 111; Conservative 76; Mismatches 147; Indels 110; Gaps

QY 144 LTSLFPVGVLLGSFGISGQLSDRFGRK-----NVLEVTMGMTQSGFLQIFSKNEFMVF 197
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 68 LSVSIFAIVGMIGSFVLGXIGNLGRKKXAMLVNVLAIAGLLMG---LAKXAASFEMLI 124
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY 198 VLFVLVGM-----GOISNVAAVFLVGLTEILGKSVRIIFFSTLGVCIFYAFGYM 244
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 125 LGRFIIGLYCGLSSGVPMYVGISEPTALRGALGT-----LNOLGIVIGILIA-Q 173
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY 245 VLPLFAYFIIRD--WRMLLVALTMPGVLCVALWNPFDIESPRWLI----SQGRFEAEVIIR 298
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 174 VLGLDLSLGNESLWPLLGLTGVPAILOLLLPFCSPRYLLINKNEARAKKALQR 238
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY 299 KAANGAVVPSTIFTDFSELQD---LSSKKQOSHNLDLRTWNTRWTVIMSIMLWMT- 353
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 234 GTADVSEV-----AEMKDSRXMKSEKXVS--VLEFRSRXYRQPVIIAIVQLSQ 283
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY 354 ----ISVGYFGLSLDTPNLHGDIFVNCFTLSAMEVEPAY-----VLAWLLLO 395
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 284 QLSGINAVEFYTSI-----FEKAGVGQPVYATIGAGVVNTVTVSVFVFE 330
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY 396 YLPRYSMATALFLGG-----SVLLFMQLVPPDLYYLATVLVMVKKECVTAAPS-----MV 446
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 331 RAGR--TLHLGLGGMACCAVLMTIALLDQVPWMXSYVISVAIFGFVAFFEVEGPGPIP 388
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY 447 VYVYTAELVPTVRNMNGVGSV---TASRLGISLSPYFV-YLGAYDRFLPYILMGSITIL 501
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 389 WFIVAELSQQPRAATAVACFSNWTSNFVGLGLFOYIAELLGPY----VFIVFAVLILL 444
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY 502 TAILT-LFLPESTGTPIDTDQM 524
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 445 FFIFTFLKVPETKGR---TFDEI 464
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

RESULT 15
US-09-299-549-10
; Sequence 10, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10

```

Search completed: January 25, 2002, 10:10:18  
Job time: 114 sec

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows95  
 SOFTWARE: FastSEQ for Windows Ver  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/299,549  
 FILING DATE: 26-APR-1999  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 09/031,392  
 FILING DATE: 26-FEB-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Weiklejohn, Ph.D., Anita L.  
 REGISTRATION NUMBER: 35,283  
 REFERENCE/DOCKET NUMBER: 07334/07  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 493 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-299-549-10

Query Match	7.08;	Score 203;	DB 4;	Length 493;
Best Local Similarity	25.09;	Pred. No. 7.3e-13;		
Matches 111;	Conservative	76;	Mismatches 147;	Indels 110; Gaps 21;
QY 144	LTISLFFGVGLGFSIGQLSDRFGRK-----NVLFVTMGQMOTGFSFLQFSKNFEMFV	197		
Db 68	LSVSIFAVGGMIGSFLVCXIGNRLGRKXAMLVNVLATAGGLMG-----LAXXASFEMLI	124		
QY 198	VLFLVGM-----GQISNVYAAFLVGLTEILGKSVRIFFSTLGVCIYAFGYM	244		
Db 125	LGRITIGLYCOLSSGVVPMVYGEISPTALRGALGT-----LNOLGTVIGILIA-Q	173		
QY 245	VLPLFAYFIRD--WRMLLVALTMFGVLCVALWMTIPESPRLI-----SOGFEBAEVLIR	298		
Db 174	VLGLDSLLGNESLWPLLLGLTGVPALLQLLLPCPESPRLYLLINKKEARAKALQRLR	233		
QY 299	KAAKANGIVVPSTIPDSELOD-----LSSKKQOQSHNILLRTWNIRMTVIMSIMLWMT-	353		
Db 234	GTADVSGEV-----AEMKDESRXMXSEKXVS--VLELFRSRXYRQPVIATVLQLSQ	283		
QY 354	-----ISGVFGLSLDTPNLHGDFIVNCFLSAMVEVPAY-----VLAWLLLQ	395		
Db 284	QLSGINAVFYYSI-----PEKAGVGQPYATIGAGVNTVFTVVSVEFVE	330		
QY 396	YLPRLYSNATALFLGG-----SVLLFMOLVPPDLYLATVLVMVGKFGVTAAFS-----MV	446		
Db 331	RAGRR--TLHLLGLGGMAGCAVLMTALALLDQVPMNSYVSIIVAFGVFAFFVGPGPIP	388		
QY 447	YVYTAELYPVWRNMGVGS-----TASRLGSLSPYFV-YLGAYDRFLPYILMGSTIL	501		
Db 389	WFIVAEFLPSGPPRAATAVAGFSNWTSNFVGLLQFIALLGPY-----VFIVFAVLLLL	444		
QY 502	TAIL-LFLPESFGTLPDPIDQM	524		
Db 445	FFITFLKVPETKGF-----TFDEI	464		

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 10:09:47 ; Search time 28.31 Seconds  
(without alignments)  
1498.736 Million cell updates/sec

Title: US-09-521-195-3  
Perfect score: 2883  
Sequence: 1 MRDYDEVTAFLGEGWPFQRL.....HTRMLKDGQERPTILKSTAF 557

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2883	100.0	557	2 JW0089	organic cation tra
2	2496	86.6	557	2 JE0346	high-affinity carn
3	777	27.0	576	2 T22509	hypothetical prote
4	733	25.4	593	2 JC4884	organic cation tra
5	712	24.7	556	2 S50862	organic cation tra
6	589	20.4	794	2 T27870	hypothetical prote
7	547.5	19.0	745	2 T16565	hypothetical prote
8	522	18.1	527	2 T01019	transport protein
9	483	16.8	539	2 C96758	probable protein
10	470.5	16.3	518	2 B86299	hypothetical prote
11	411	14.3	528	2 T21682	hypothetical prote
12	382.5	13.3	540	2 T25851	hypothetical prote
13	380.5	13.2	515	2 B96825	hypothetical prote
14	368.5	12.8	521	2 H86298	hypothetical prote
15	355.5	12.3	529	2 T23190	hypothetical prote
16	341.5	11.8	454	2 F75380	probable sugar tra
17	316	11.0	455	2 B83213	probable MFS trans
18	308	10.7	469	2 G65058	hypothetical prote
19	298.5	10.4	724	2 I50531	transmembrane tran
20	295.5	10.2	423	2 S74046	probable sugar tra
21	291.5	10.1	435	2 T15290	hypothetical prote
22	287	10.0	400	2 C69757	transporter homolo
23	287	10.0	422	2 G72234	hypothetical prote
24	279	9.7	524	2 T27082	hypothetical prote
25	274	9.5	683	2 S34961	synaptic vesicle p
26	272	9.4	478	2 T33985	hypothetical prote
27	271	9.4	480	2 T23608	hypothetical prote
28	267	9.3	443	2 E64725	yaau protein - Esc
29	267	9.3	443	2 H85485	probable transport

30	266	9.2	423	2 T19030	hypothetical prote
31	265	9.2	459	2 G64937	hypothetical prote
32	265	9.2	459	2 H85787	probable transport
33	262	9.1	442	2 A83122	probable MFS trans
34	261.5	9.1	742	1 A43344	synaptic vesicle p
35	258	8.9	742	2 S27263	metabolite transpo
36	256	8.9	482	2 B69803	hypothetical metab
37	255.5	8.9	452	2 A64937	hypothetical prote
38	253	8.8	520	2 T23545	arabinose-proton s
39	250.5	8.7	472	2 S47089	hypothetical prote
40	250.5	8.7	517	2 T19962	sugar transporter
41	249.5	8.7	401	2 E69501	probable transport
42	248	8.6	416	2 H85786	metabolite transpo
43	243	8.4	461	2 D70073	hypothetical prote
44	242	8.4	312	2 T25852	hypothetical prote
45	241.5	8.4	457	2 E70070	metabolite transpo

## ALIGNMENTS

RESULT 1  
JW0089  
organic cation transporter protein 2 - human  
N:Alternate names: OCTN2  
C:Species: Homo sapiens (man)  
C:Date: 18-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 28-May-1999  
C:Accession: JW0089  
R:Wu, X.; Prasad, P.D.; Leibach, F.H.; Ganapathy, V.  
Biochem. Biophys. Res. Commun. 246, 589-595, 1998  
A:Title: cDNA sequence, transport function, and genomic organization of human OCTN2,  
A:Reference number: JW0089; MUID:98289574  
A:Accession: JW0089  
A:Molecule type: mRNA  
A:Residues: 1-557 <WUA>  
A:Cross-references: GB:AF057164; NID:g3273740; PIDN:AAC24828.1; PID:g3273741  
A:Experimental source: placenta  
C:Comment: This transporter functions in the elimination of cationic drugs and other

Query Match	100.0%;	Score 2883;	DB 2;	Length 557;
Best Local Similarity	100.0%;	Pred. No. 7.8e-216;		
Matches 557;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MRDYDEVTAFLGEGWPFQRLIFLLSASII	PNGFTGLSSVFLIATPEHRCRVPDAA	LS 60
Db	1	MRDYDEVTAFLGEGWPFQRLIFLLSASII	PNGFTGLSSVFLIATPEHRCRVPDAA	LS 60
QY	61	AWRNHTVPLRLRDGRVPHSCRRYRLATIANF	SALGLEPGRVDLGLQEQESCLDGEW	FE 120
Db	61	AWRNHTVPLRLRDGRVPHSCRRYRLATIANF	SALGLEPGRVDLGLQEQESCLDGEW	FE 120
QY	121	QDVYLSITVTENLVGCEDDWKAPLTISLFFV	GVLLGSFTSGQLSDRFRGNVLFVTMG	MQ 180
Db	121	QDVYLSITVTENLVGCEDDWKAPLTISLFFV	GVLLGSFTSGQLSDRFRGNVLFVTMG	MQ 180
QY	181	TGFSFLQIFSKNFEMVFLVVGWGOISNYAA	FVLGVILGKSVRIIFSLGVCIFYA	240
Db	181	TGFSFLQIFSKNFEMVFLVVGWGOISNYAA	FVLGVILGKSVRIIFSLGVCIFYA	240
QY	241	FGYMWLPFLFAYFTRDWRMLLVALTMPGVLC	VALWVFIPESPRWLISQGRFEAEVILR	KA 300
Db	241	FGYMWLPFLFAYFTRDWRMLLVALTMPGVLC	VALWVFIPESPRWLISQGRFEAEVILR	KA 300
QY	301	AKANGIVVPESTIFDPSELQDLSSKKQOSH	NIIDLRTWNIRMTIMSLMWTISVGY	FG 360
Db	301	AKANGIVVPESTIFDPSELQDLSSKKQOSH	NIIDLRTWNIRMTIMSLMWTISVGY	FG 360
QY	361	LSLDTPNLHGDIFVNCFLSAMVEVPAYVLA	WLLQLYLPRIYSMATALFGLGSGVLL	FQMQLV 420
Db	361	LSLDTPNLHGDIFVNCFLSAMVEVPAYVLA	WLLQLYLPRIYSMATALFGLGSGVLL	FQMQLV 420
QY	421	PPDLYYLATVLVVGKFGVTAAFSMVYVTA	EELYPTVVRNMGVGVSSSTASRLG	SILSPYF 480

```

Db 421 PDLYLATLVVMGKFGVTAFAFVYVYTAELVTVVNMGVGVSSVTSASRLSPYF 480
QY 481 VYLGAYDRFLPYLMGSLTILTAITLFLPSFGPLPDTIDQMLRVKGMKHKRTPSHR 540
Db 481 VYLGAYDRFLPYLMGSLTILTAITLFLPSFGPLPDTIDQMLRVKGMKHKRTPSHR 540
QY 541 MLKDGQERTILKSTAF 557
Db 541 MLKDGQERTILKSTAF 557

RESULT 2
JF0346
high-affinity carnitine transporter, Cnl - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JF0346
R:Seikine, T.; Kusuhabara, H.; Utsunomiya-Tate, N.; Tsuda, M.; Sugiyama, Y.; Kanai, Y.; Endo
Biochem. Biophys. Res. Commun. 251, 586-591, 1998
A:Title: Molecular cloning and characterization of high-affinity carnitine transporter
A:Reference number: JF0346; MUID:99011422
A:Accession: JF0346
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-557 <SEK>
A:Cross-references: DDBJ:AB017260; NID:g3869208; PID:BAA34399.1; PID:g3869209

Query Match 86.6%; Score 2496; DB 2; Length 557;
Best Local Similarity 85.3%; Pred. No. 7.8e-186;
Matches 475; Conservative 42; Mismatches 40; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEMGFQRLIFLLS-ASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
Db 1 MRDYDEVTAFLGEMGFQRLIFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
QY 61 AWRNHTVPLRLRDREVPCHSCRRYRLATIANFSALGLEPGRDVLGQLEQESCDLGNWFS 120
Db 61 AWRNHSIPLTKDGRQVPQSCRRYRLATIANFSALGLEPGRDVLGQLEQESCDLGNWFS 120
QY 121 QDVYSTIVTEENLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRGKKNVFLVTGMQ 180
Db 121 QDVYSTIVTEENLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRGKKNVFLVTGMQ 180
QY 181 TGFSLQIFSKNFEMFVFLVGVGQISNYVAAFVLTGTEILGKSVRIIFSTLGCIFYA 240
Db 181 TGFSLQLFSVNFETVFLVGVGQISNYVAAFVLTGTEILGKSVRIIFATLGCIFYA 240
QY 241 FGYVLPFLFAFIRDRMMLVALTMPGVLCVALMWFTPESPRWLISQGRPEAEVIRKA 300
Db 241 FGYVLPFLFAFIRDRMMLVALTMPGVLCVALMWFTPESPRWLISQGRPEAEVIRKA 300
QY 301 AKANGIVVPSTIFDPSELODLSKKQ--OSHNIIDLLRTNIRMTIMSLMWTISVGYG 360
Db 301 AKANGIVVPSTIFDPSELODLSKKQ--OSHNIIDLLRTNIRMTIMSLMWTISVGYG 360
QY 361 LSLDTPNLHGDIFFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLFQ 420
Db 361 LSLDTPNLHGDIFFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLFQ 420
QY 421 PPDLYYLATLVVMGKFGVTAFAFVYVYTAELVTVVNMGVGVSSVTSASRLSPYF 480
Db 421 PSELYLATLVVMGKFGVTAFAFVYVYTAELVTVVNMGVGVSSVTSASRLSPYF 480
QY 481 VYLGAYDRFLPYLMGSLTILTAITLFLPSFGPLPDTIDQMLRVKGMKHKRTPSHR 540
Db 481 VYLGAYDRFLPYLMGSLTILTAITLFLPSFGPLPDTIDQMLRVKGMKHKRTPSHR 540
QY 541 MLKDGQERTILKSTAF 557
Db 541 TOKDGGESPTVLKSTAF 557

```

```

RESULT 3
T22509
hypotheical protein F52F12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22509
R:Matthews, L.
submitted to the EMBL Data Library, December 1996
A:Reference number: Z19573
A:Accession: T22509
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-576 <WIL>
A:Cross-references: EMBL:283228; PIDN:CA05732.1; GSPDB:GN00019; CESP:F52F12.1
A:Experimental source: clone F52F12
C:Genetics:
A:Gene: CESP:F52F12.1
A:Map position: 1
A:Introns: 16/1; 23/3; 50/1; 80/3; 108/2; 134/3; 221/1; 257/2; 319/3; 358/2; 490/3; 5

Query Match 27.0%; Score 777; DB 2; Length 576;
Best Local Similarity 31.4%; Pred. No. 1.5e-52;
Matches 165; Conservative 118; Mismatches 225; Indels 18; Gaps 6;

QY 3 DYDEVTAFLGEMGFQRLIFLLS-ASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 61
Db 16 DFDVLEQVNGYQIVVFFIICLPTSLPSAFSAFNIFVVGNNPHTCHIEGKEYLAP 75
QY 62 WRNHTVPLRLRDREVPCHSCRRYRLATIANFSALGLEPGRDVLGQLEQESCDLGNWFS 121
Db 76 LTNDTQIL-----SCQYNETQINVFRAFTSAP-VDTYSDRISLVPQNGWDYDN 124
QY 122 DVYSTIVTEENLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRGKKNVFLVTGMQ 181
Db 125 STYLSLVTEENLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRGKKNVFLVTGMQ 184
QY 182 GFSFLQIFSKNFEMFVFLVGVGQISNYVAAFVLTGTEILGKSVRIIFSTLGCIFYA 241
Db 185 VGTASSFAKDIESFIILRFTGLAFLPFIIFCMEFMGNSGR-IFSGLWTSIFFGA 243
QY 242 GYVLPFLFAFIRDRMMLVALTMPGVLCVALMWFTPESPRWLISQGRPEAEVIRKA 301
Db 244 AMALLGVAMFIRWRQLTFFCNAPFAFVYIYFFFLPSRWSVSVGWADAKKQK 303
QY 302 KANG--IVVPSTIFDPSELODLSKKQ--OSHNIIDLLRTNIRMTIMSLMWTISV 356
Db 304 KMGKSNVDVDELVDMSKKNHQAEEKTKRSHNVTDLFTPLNRKTLIVYIWMNAI 363
QY 357 GYFGLSDTPNLHGDIFFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLF 416
Db 364 IYNGTLINSLNPVDDYWSFIINGAVELPGYFVFWPLOCAGRRWTAAATMIVCGIGCVS 423
QY 417 MQLVPPDLYYLATLVVMGKFGVTAFAFVYVYTAELVTVVNMGVGVSSVTSASRLS 476
Db 424 AMFMPDGYPLWASAFIGKFGVSGFAVIYIFAGELYPTVVRIGMGMSSVAGSGLL 483
QY 477 SPYFVYLVGAYDRFLPYLMGSLTILTAITLFLPSFGPLPDTID 522
Db 484 APHIVNLGKIVKILPLIMGLMALSAGIITFFLPETLGAFLPMTIE 529

RESULT 4
JC4884
organic cation transporter protein 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
C:Accession: JC4884
R:Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.
Biochem. Biophys. Res. Commun. 224, 500-507, 1996
A:Title: cDNA cloning and functional expression of a novel rat kidney organic cation
A:Reference number: JC4884; MUID:96295517

```



A:Accession: JC4884

A:Molecule type: mRNA

A:Residues: 1-593 &lt;ORU&gt;

A:CROSS-references: DDBJ:DB3044; NID:g1502282; PIDN:BA11754.1; PID:d1012421; PID:g15022

A:Experimental source: kidney

C:Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 25.4%; Score 733; DB 2; Length 593;  
Best Local Similarity 35.2%; Pred. No. 3.9e-49;  
Matches 198; Conservative 89; Mismatches 234; Indels 42; Gaps 15;

QY 1 MRDYDEVTAFLEWGPQRLIFLLSASIPNGFTGL--SSVPLIATPEHRCRVPDAANL 58

Db 1 MSTVDDILEHIGFHLFKQKTFELL--ALLSGAFTPIYGVFLGTPDHCWSPGAAL 58

QY 59 SS---AWR-----NHTVP-LRLRDCREVPHSCRRYR-----LATIANFSALGLEPGRDV 103

Db 59 SQRGWSQAEELNVTVPGLGPSDEASFLSQCMRYEVDWQSTLDCVDPSSLA-----A 112

QY 104 DLGOEESCLDGEWFSQDYLSTIVTEWNLVCEDDWKAPLTLSLFFVGLGFSFISGOL 163

Db 113 DRNGLPLGCPGHEGVNTPG--SSIVTEFNLCVCAHSMWMLDLFQSVVNVGFFIGAMMIGYL 170

QY 164 SDRGRKNVLFTVMGQTGFSLQIFSKNFMFVFLVFLVGMGQISNYVAAFVLGTEILG 223

Db 171 ADPRGRKFCLLVTLINAIISGALMAISPNAVWMLVRFELQGLVSKAGWLGILITLIFVYG 230

QY 224 KSVRIIFSTLGVCIYAF--GYMVLPLFAFYFIRDRMLLVALTMPGVLCVALWETPESP 281

Db 231 LGYR---RMVGICYQTAFTVGLLILAGVAVIPNWRLOFAVTLNFCFLYFWCIPESP 287

QY 282 RWLSQGRFEAEVIRKAKANGIVVPSTIFDPSELQDLSKKQKSHNLDLRLTWNIR 341

Db 288 RWLSQNKIVKAMKIHKIAKNGKSPVSLQNLTPDEDAGKCLKPS--ILLIVRTPQIR 345

QY 342 MVTIMSILMANTISVGVFGLSLDTPNLHGD-IFVNCFLSAMVEVPAYVLAWLLQLVLP 400

Db 346 KHTLILMNYNFTSSVLQGGILMIM-GLAGDNIYLDFFYSALVEFPAAFTIITIDRVGR 404

QY 401 YSMATALLFGSVLLFMQLVPPDLYLATVLMVVGKFGVTAAPSMVYVTAELYPTVVRN 460

Db 405 YPWAIVSNWAGAACLASVFPDQLKTIACLRGMITMAYEMVCIYNAELYPTIYIRN 464

QY 461 MGVGVSSTARLGSILSPFYV-LGAYDRFLPYILMGSLTILTAITLFLPESFGPLPD 519

Db 465 LGVLVSSMCDIGGIITPFLVYRLTDIWMFEPLVVFVAVGLVAGALVLLLPETKKGALPE 524

QY 520 TID----QMLRVKGMKHKRTPS 537

Db 525 TIEDAENMQRPKRKERNLPPS 547

RESULT 5

S50862

organic cation transport protein Ocr1 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 05-Nov-1999

C:Accession: S50862; S78533; I58089

R:Gruendemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.

Nature 372, 549-552, 1994

A:Title: Drug excretion mediated by a new prototype of polyspecific transporter.

A:Reference number: I58089; MUID:95082907

A:Accession: S50862

A:Molecule type: mRNA

A:Residues: 1-556 &lt;GRU&gt;

A:CROSS-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622

A:Experimental source: kidney

R:Gorboulev, V.G.

submitted to the EMBL Data Library, January 1995

A:Reference number: S78533

A:Accession: S78533

A:Molecule type: mRNA

A:Residues: 1-342, 'N', 344-556 &lt;GOR&gt;

A:CROSS-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622

C:Keywords: glycoprotein; phosphoprotein; transmembrane protein

F:20-46/Domain: transmembrane #status predicted &lt;TM1&gt;

F:154-171/Domain: transmembrane #status predicted &lt;TM2&gt;

F:178-197/Domain: transmembrane #status predicted &lt;TM3&gt;

F:243-260/Domain: transmembrane #status predicted &lt;TM4&gt;

F:267-283/Domain: transmembrane #status predicted &lt;TM5&gt;

F:350-366/Domain: transmembrane #status predicted &lt;TM6&gt;

F:380-398/Domain: transmembrane #status predicted &lt;TM7&gt;

F:406-425/Domain: transmembrane #status predicted &lt;TM8&gt;

F:435-452/Domain: transmembrane #status predicted &lt;TM9&gt;

F:469-485/Domain: transmembrane #status predicted &lt;TM10&gt;

F:494-514/Domain: transmembrane #status predicted &lt;TM11&gt;

F:71.97.113.432/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:286.292/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pred

F:296.343.550/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status

Query Match 24.7%; Score 712; DB 2; Length 556;

Best Local Similarity 33.8%; Pred. No. 1.5e-47;

Matches 193; Conservative 96; Mismatches 230; Indels 52; Gaps 14;

QY 1 MRDYDEVTAFLEWGPQRLIFLL---SASIPNGFTGLSSVFLIATPEHRCRVPDAAN 57

Db 1 MPTVDVLEQGEFGWFOQAFLLLCLLSASLAPI-YVGI--VFLGFTPGHYCQNPQVAE 57

QY 58 LSS---AWR-----NHTVP-LRLRDCREVPHSCRRYRL-----ATTANFSAL 95

Db 58 LSQRGWSQAEELNVTVPGLGPSDEASFLSQCMRYEVDWQSTLDCVDPSSLVANKSQL 117

QY 96 GLEPGRDVLGOLESCLDGEWFSQDYLSTIVTEWNLVCEDDWKAPLTLSLFFVGVLL 155

Db 118 PLGP-----CEHGMY--DTPGSSIVTEFNLCVCAHSMWMLDLFQSVVNVGFFL 162

QY 156 GSFTSGQLSDRGRKNVLFTVMGQTGFSLQIFSKNFMFVFLVFLVGMGQISNYVAAF 215

Db 163 GSLVYGVYIADRGRKLCVLTTLVTSVSGVLTAVAPDYSMLLFRLLQGMVSKGWSVSGY 222

QY 216 VLGTIELKSKVRIIFSTLGVCIYAF--GYMVLPLFAFYFIRDRMLLVALTMPGVLCVAL 273

Db 223 TLITEFVSGYR---RTTAYIQMAFTVGLVGLAGVATAIPDRMWLQVLAIVLFLFLY 279

QY 274 WMFIPESRWLISQGRFEAEVIRKAKANGIVVPSTIFDPSELQDLSKKQKSHNILD 333

Db 280 YWFVPESPRWLLSQKRTTRAVIMEQIAQNGKVPADLKMCLCEEDASER--SPSFAD 337

QY 334 LLRTWNIRMTIMSLMWTISVGVFGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLL 393

Db 338 LFRTPTLRKHHTVILMYLWFSQAVLYQGLIMHVGTAGANLYLDFFYSILVEPPAAFIILVT 397

QY 394 LOYLPRRYSMATALLFGSVLLFMQLVPPDLYLATVLMVVGKFGVTAAPSMVYVTAEL 453

Db 398 IDRTGRITYPIAASNLVTGAACLLMIFIPHEUHLNVTIACLRMGATIVLOMCLVNAEL 457

QY 454 YPTVVRNMGVGVSTASRLGSILSPFYV-LGAYDRFLPYILMGSLTILTAITLFLPLPES 512

Db 458 YPTFIRNLGMVMSALCDLGGITFPFNVFRLMEVWQALPLILFGLVGLITAGAMTLLLPET 517

QY 513 PGTPLPDTIDQMLRVKGMKHKRTPSFRMLK 543

Db 518 KGVALPETIEEAENL-GRKSKAKENTIIYQ 547

RESULT 6

T27870

hypothetical protein ZK455.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T27870

R:White, S.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z20432



A,Map position: 1

Query Match 18.1%; Score 522; DB 2; Length 527;  
Best Local Similarity 28.3%; Pred. No. 7.7e-33;  
Matches 155; Conservative 100; Mismatches 206; Indels 86; Gaps 17;  
QY 5 DEVTAFLGNGWGFQRLIFLLSASIIPIGFTGLSSVFLIATPEHRCRVPDPAANLSSAWN 64  
DB 22 DTIESYIGFQWAGFQALVFSFGVDAQQTFFISVTDSEPTWHC----- 67  
QY 65 HTVPLRLDRGVRPHSCRRYRLATIANFALGLEPGRDVDLQLEQESCLDGEWESQDVY 124  
DB 68 -----TDSNSICHE-----SISNICL-----PKTANSWDYSFHV- 97  
QY 125 LSTIVTEWNLVEDDKAPLTISLFFVGVLLGSFISGQLSD-RFGKKNVLFVTMGQTGF 183  
DB 98 --SVISEWGLQAGSFVKGLPESSEFFVGLIGGLVLTADSSLGKKNMLFLSCLVMAIS 155  
QY 184 SFLQIFSKNFEMFVFLVFLVGMGOISNYAAAFVLGTEILGKSVRIIFSTLGVCIYAF-- 241  
DB 156 TMTVFPNIWYAVLRFVNGFGRATIGCALVLTSLVGVKKWR---GRVGIMSEFGWML 212  
QY 242 GYMLVPLFAFYFIR--DWRMLLVALTMPGVL-CVALWVFIPESPRWLISQGRFEAEVIR 298  
DB 213 GFLSLPLMAYMNRGSSWRILYAWTSIPTIYCVLRFVFCESPRWLFVGRGREAISILK 272  
QY 299 KAANGIVVPST-----IFDPSELQDLSKKQOSHNLIDLLRT-----WNI-RM 342  
DB 273 RVAS-----IPSTDVSGGAISMSFSESLPFEDEEKPSTNVNIFTMTKVLVLEKRWALKRL 327  
QY 343 VTMTSMLMWTISGVFGLSLDTPNLHGDIFFVNCFLSAMVEVPAYVLAWLLOLPLPRYS 402  
DB 328 SAVMAIA--FGIGLVYGMPLANSLDNFNLYSAAFNALMDLPANLITLFLVDKLSRNA 385  
QY 403 MATALFLGG--SVLLF-----MQLVPPDLYLATVLMVGKFGVTAFAFMYVYVTELYPT 456  
DB 386 LIGFTALGGVSSVLIPTALHNMIRGNHGLQLA--LELISVFSACSAFNMEIYITTELEPT 443  
QY 457 VVRNMGVYSTASRLGSLSPYFVYLGAIDRPLPYILMGSLTILTAITLFLPESFGTP 516  
DB 444 CVRNSAIFAMARQALVLGGVFPIMVAAGRNKAFWSFGLGLAIGLGLFAVGLPETRGS 503  
QY 517 LPDTIQD 523  
DB 504 LCDTMD 510

RESULT 9  
C96758  
probable protein transporter T18K17.11 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C96758  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C96758  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-539 <STO>  
A:Cross-references: GB:AE005173; NID:96598860; PIDN:AAF18714.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 16.8%; Score 483; DB 2; Length 539;  
Best Local Similarity 29.0%; Pred. No. 8.4e-30;  
Matches 157; Conservative 90; Mismatches 209; Indels 86; Gaps 18;  
QY 5 DEVTAFLGNGWGFQRLIFLLSASIIPIGFTGLSSVFLIATPEHRCRVPDPAANLSSAWN 63  
DB 46 DEVTEQHIGALGFAQILHALLVSIANIFDAQTWLSIFSDAQP-----AARL----- 92  
QY 64 NHTVPLRLDRGVRPHSCRRYRLATIANFALGLEPGRDVDLQLEQESCLDGEWESQDVY 120  
DB 93 -----LATG-----AIVEGASLCGLASGE-----WEMIGPK 118  
QY 121 QDVYLSITVTEWNLVEDDKAPLTISLFFVGVLLGSFISGQLSDR-FGRKKNVLFVTMG 179  
DB 119 SD-----TVSEWNLICQHKFLVAVPSTLFIPLGSLFGSGVGYLADSWFGRRKTLILLSCVL 174  
QY 180 QTFSEFLQIFSKNFEMFVFL-----FVLVGMGOISNYAAAFVLGTEILGKSVRIIFSTLG 234  
DB 175 TFTAFAISFSPNVMVYAFRLFANGFFRSIGTS-----CCIVLATEIVGKKRGQVGY 229  
QY 235 VCIFYAFGYWVLPFAFYFIR--DWRMLLVALTMPGVL-CVALWVFIPESPRWLISQGRFE 292  
DB 230 F-FFFTLGFSLPLMAYLERSKRNLYRIISFLPLGAVCLLPAYESPRWLVKGRNKE 288  
QY 293 AEVIRKAANKANGIVVPS--TIFDP-SELQDLSKKQOSHNLIDLLRTWNIWMTMSIM 349  
DB 289 AMVVLKRLARLNGQLPADLSLVDPIPERDDOTSSEKFWK-----TKWAVKRI-INVM 342  
QY 350 LMWTISGVYFGLSLDTPNLHGDIFFVNCFLSAMVEVPAYVLAWLLOLPLPRYSNATAFL 409  
DB 343 AGFGSGFVYIGLNAENLNENLYTVAVNALMEFPAPVAFSGFLGVMNRPLFSNSYL 402  
QY 410 GGSVLLFMQLVP-----PDLYLATVLMVGKFGVTAFAFMYVYVTELYPTVVRNM 461  
DB 403 AGFACLCLCAVLSIHRVIRASVAKVQLAVEAGFMASSTAYDVLVYICVELFPTNVRT 462  
QY 462 GGVYSTASRLGSLSPYFVYLGAIDRPLPYILMGSLTILTAITLFLPESFGTPPLDPTI 521  
DB 463 AVSLLRQAFMLGASAPLLVALGRESAMMSFIVGVASVLSIVSLMLRETRNAPLYETL 522  
QY 522 DQ 523  
DB 523 AQ 524

RESULT 10  
B86299  
hypothetical protein AAD34691.1 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: B86299  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: B86299  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-518 <STO>  
A:Cross-references: GB:AE005172; NID:94966360; PIDN:AAD34691.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

[illegible]



submitted to the EMBL Data Library, April 1993  
A:Reference number: Z19704  
A:Accession: T23190  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-529 <WIL>  
A:Cross-references: EMBL:Z22175; PIDN:CAA80131.1; GSPDB:GN00021; CESP:ZK637.1  
A:Experimental source: Clone K01F9  
C:Genetics:  
A:Gene: CESP:ZK637.1  
A:Map position: 3  
A:Introns: 31/2; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3  
C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match	12.3%	Score 355.5;	DB 2;	Length 529;
Best Local Similarity	26.5%	Pred. No. 6.3e-20;		
Matches 140;	Conservative 97;	Mismatches 186;	Indels 105;	Gaps 24;

QY	52	VPOAANLSSAARNHVTPLRLRDGRVPHSCRRYRLATIANFSALGLEP--GRDVD-----	104
DB	10	VLEASNTEAYVDLTAKQLIKEIRHVGDDE-AVRYNLDORTELG-EPTDQRSPPDSEKTF	67
QY	105	-----LGQLQOE-SCLDGWFSQD-----VYLSTI-----VTEWNLVCEDDWKAPLT	145
DB	68	TVDEAVEALGFGFQKLSILTCGMANWADAMENMLLSLSPALACEWGI---SSVQQALV	124
QY	146	ISLFFVGVLLGSFISQSDREG-RKNVLEVT-----MGMTQGS---FLQIFSKNFEMP	196
DB	125	TTCVFSGMMLSTFWKICDQFGRKGLTFSTLVACIMGVISGMSPHFYLLFFRGLTGF	184
QY	197	VVLFVLVGMGOISNYAAV--LGTEILGKSVRIIFSTLGVCFYAFGYMVLPLFAYFIR	254
DB	185	-----GIGGVPSVTLVAEFLPTAQRACVVLIES-----FWAIGAVFEALLAYFVM	231
QY	255	D---WRMLLVALTMP-CVLCVALWTFIPESPRWLISQGRFEAEVIRKAANGIVVPS	310
DB	232	ESTGWRALMFLSLPLIGIFAVASFW-LPESARFDMASGHPERALETQAAARNRVQLPT	290
QY	311	TIPDPSELQDLSSKKQOOSHNLDL--LRTWNIRWVTIMSIMLWMTISVGYFGLSLDTPNL	368
DB	291	G-----RLVSTKAGSESRGDIANLLSPDLRKTKTILLWCIAWATAFSYIGWVLTIVL	343
QY	369	-----HGDIFVN-----CFLSAMVEVPAYVLAWLLQLYLPRRYSMA	404
DB	344	FQSHDECHGGLFSNGTQMEVCOPLTRSDYFDLLSTTLAEFPGLIITVLIIEWFGKTKMA	403
QY	405	T--ALFLGGSVLLFQVLPDLYLATVLMYKFGVTAAFSMVVYVYTAELYPTVVRNMG	462
DB	404	LEVAVPAITFLLYICLD----RFTVTVLIFVARAFISGAFOCAYVYTPVYPTTLRAVG	459
QY	463	VGVSTASRLGSLGPYFYVLGAYDRFLPYILMGSLTITAILTLFLP	510
DB	460	LGTCSAMARIGAIVA-----SEKSLSLPIGIYGTAILGLIASLSLP	501

Search completed: January 25, 2002, 10:09:48  
Job time: 84 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 10:15:19 ; Search time 18.96 Seconds

(without alignments)  
1077.126 Million cell updates/sec

Title: US-09-521-195-3

Perfect score: 2883

Sequence: 1 MRDYDEVTAFLGEGWPFQRL.....HTRMLKDGQERTILKSTAF 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2883	100.0	557	1 OCN2_HUMAN	O76082 homo sapien
2	2513	87.2	557	1 OCN2_MOUSE	Q920e8 mus musculus
3	2509	87.0	557	1 OCN2_RAT	P46501 rattus norv
4	387.5	13.4	751	1 YLX5_CAEEL	P46501 caenorhabdi
5	357	12.4	1222	1 YMP3_CAEEL	Q10947 caenorhabdi
6	355.5	12.3	529	1 YOL1_CAEEL	P30638 caenorhabdi
7	308	10.7	445	1 YGC5_ECOLI	Q46909 escherichia
8	291.5	10.1	435	1 YTI3_CAEEL	Q10917 caenorhabdi
9	287	10.0	400	1 YCEI_BACSU	O34691 bacillus su
10	267	9.3	443	1 YAAU_ECOLI	P31679 escherichia
11	265	9.2	459	1 YDJK_ECOLI	P76230 escherichia
12	262.5	9.1	742	1 SYV2_RAT	Q02563 rattus norv
13	256	8.9	482	1 YFIG_BACSU	P54723 bacillus su
14	255.5	8.9	452	1 YDJE_ECOLI	P38055 escherichia
15	252.5	8.8	457	1 PCAK_ACTICA	Q43975 acinetobact
16	250.5	8.7	472	1 ARAE_KLEOX	P45598 klebsiella
17	243.5	8.4	551	1 HGT1_KLULA	P49374 kluyveromyc
18	241.5	8.4	472	1 ARAE_ECOLI	P09830 escherichia
19	240.5	8.3	495	1 GTR3_CANFA	P47842 canis famli
20	239.5	8.3	451	1 YFAJ_BACSU	P37514 bacillus su
21	239	8.3	461	1 CSBC_BACSU	P46333 bacillus su
22	237.5	8.2	496	1 GTR3_HUMAN	P11169 homo sapien
23	234	8.1	494	1 GTR3_SHEEP	P47843 ovis aries
24	227	7.9	490	1 GTRL_CHICK	P46896 gallus gall
25	226.5	7.9	484	1 GALP_ECOLI	P37021 escherichia
26	224.5	7.8	413	1 MUCK_ACTICA	P94131 acinetobact
27	223	7.7	451	1 GTRL_PIG	P20303 sus scrofa
28	222	7.7	763	1 RGT2_YEAST	Q12300 saccharomyc
29	221	7.7	466	1 BENK_ACTICA	O30513 acinetobact
30	221	7.7	492	1 GTRL_RAT	P11167 rattus norv
31	220	7.6	492	1 GTRL_BOVIN	P27674 bos taurus
32	219.5	7.6	491	1 XYLE_ECOLI	P09098 escherichia
33	219	7.6	493	1 GTR3_MOUSE	P32037 mus musculus

## RESULT 1

ID	OCN2_HUMAN	STANDARD;	PRT;	557 AA.
AC	O76082;			
DT	20-AUG-2001	(Rel. 40, Created)		
DT	20-AUG-2001	(Rel. 40, Last sequence update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22, MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).			
DE	SLC22A5 OR OCTN2			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9828574; PubMed=9618255;			
RA	Wu X., Prasad P.D., Leibach F.H., Ganapathy V.;			
RT	"cDNA sequence, transport function, and genomic organization of human OCTN2, a new member of the organic cation transporter family.";			
RL	J. Biochem. Biophys. Res. Commun. 246:589-595(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE=Kidney;			
RC	MEDLINE=98352077; PubMed=9685390;			
RA	Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M., Sai Y., Tsuji A.;			
RT	"Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2.";			
RL	J. Biol. Chem. 273:20378-20382(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99113835; PubMed=9916797;			
RA	Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N., Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsushita T., Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M., Tsuji A.;			
RT	"Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter.";			
RL	Nat. Genet. 21:91-94(1999).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=99384224; PubMed=10454528;			
RA	Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H., Chen J., Conway S.J., Ganapathy V.;			
RT	"Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter.";			
RL	J. Pharmacol. Exp. Ther. 290:1482-1492(1999).			
RN	[5]			
RP	VARIANT CDSP GLN-169.			
RX	MEDLINE=99355597; PubMed=10425211;			
RA	Burwinkel B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K., Gerbitz K.-D., Kilian M.W.;			
RT	"Carnitine transporter OCTN2 mutations in systemic primary carnitine deficiency: a novel Arg169Gln mutation and a recurrent Arg282ter			

P17809 mus musculus  
P28568 gallus gall  
P11166 homo sapien  
P13355 oryctolagus  
P53142 saccharomyc  
Q51955 pseudomonas  
P10870 saccharomyc  
Q07647 rattus norv  
P40885 saccharomyc  
P54862 saccharomyc  
Q01440 leishmania  
P38695 saccharomyc

## ALIGNMENTS

RT mutation associated with an unconventional splicing abnormality.";

RL Biochem. Biophys. Res. Commun. 261:484-487(1999).

[6]

RP VARIANT CDSP CYS-211.

RX MEDLINE=99408248; PubMed=10480371;

RA Vaz F.M., Scholte H.R., Ruiter J., Hussaarts-Odijk L.M.,

RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,

RA Wanders R.J.A.;

RT "Identification of two novel mutations in OCTN2 of three patients with

RT systemic carnitine deficiency.";

RL Hum. Genet. 105:157-161(1999).

[7]

RN VARIANT CDSP LEU-478.

RX MEDLINE=99172075; PubMed=10072434;

RA Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,

RA Wanders R.J., Fok T.F., Hjeltn M.M.;

RT "Mutations of OCTN2, an organic cation/carnitine transporter, lead to

RT deficient cellular carnitine uptake in primary carnitine deficiency.";

RL Hum. Mol. Genet. 8:655-660(1999).

[8]

RN CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.

RX MEDLINE=20026865; PubMed=10559218;

RA Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;

RT "Mutations in novel organic cation transporter (OCTN2), an organic

RT cation/carnitine transporter, with differential effects on the

RT organic cation transport function and the carnitine transport

RT function.";

RL J. Biol. Chem. 274:33388-33392(1999).

[9]

RN VARIANTS CDSP ARG-283 AND PHE-446.

RX MEDLINE=20081068; PubMed=10612840;

RA Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,

RA Tsuji A.;

RT "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a

RT patient with primary systemic carnitine deficiency.";

RL Hum. Mutat. 15:118-118(2000).

[10]

RN VARIANT CDSP LYS-452.

RX MEDLINE=20145665; PubMed=10679939;

RA Wang Y., Kelly M.A., Cowan T.M., Longo N.;

RT "A missense mutation in the OCTN2 gene associated with residual

RT carnitine transport activity.";

RL Hum. Mutat. 15:238-245(2000).

CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE

CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE

CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF

CC CARNITINE.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,

CC HEART AND PLACENTA.

CC -1- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY

CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE

CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE

CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND

CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL

CC MYOPATHY OR CARDIOMYOPATHY.

CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC

CC CATION SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

CC -----

DR EMBL: AF057164; AAC24828.1; -

DR EMBL: AB015050; BAA29023.1; -

DR EMBL: AB016625; BAA36712.1; -

DR MIN: G03377; -

DR MIN: 212140; -

DR InterPro: IPR003662; sub\_transportr.

DR Pfam: PF00083; sugar\_tr; 1.

DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.

KW Transport; Transmembrane; Glycoprotein; Disease mutation.

FT TRANSMEM 21 41

FT TRANSMEM 143 163

FT TRANSMEM 173 193

FT TRANSMEM 196 216

FT TRANSMEM 233 253

FT TRANSMEM 258 278

FT TRANSMEM 343 363

FT TRANSMEM 372 392

FT TRANSMEM 414 434

FT TRANSMEM 437 457

FT TRANSMEM 489 509

FT CARBOHYD 57 57

FT CARBOHYD 64 64

FT CARBOHYD 91 91

FT VARIANT 169 169

FT VARIANT 211 211

FT VARIANT 283 283

FT VARIANT 446 446

FT VARIANT 452 452

FT VARIANT 478 478

FT TRANSPORT BUT STIMULATED ORGANIC CATION

FT TRANSPORT.

FT /FTID=VAR\_009257.

FT M->R: LOSS OF BOTH CARNITINE AND ORGANIC

FT CATION TRANSPORT FUNCTIONALITIES.

FT SQ SEQUENCE 557 AA; 62751 MW; 928B1F6EFF63C48D CRC64;

Query Match 100.0%; Score 2883; DB 1; Length 557;

Best Local Similarity 100.0%; Pred. No. 1.8e-178;

Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEGWGPQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDAAANLSS 60

DB 1 MRDYDEVTAFLGEGWGPQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDAAANLSS 60

QY 61 AWRNHTVPLRLRGREVPHSCRRYRLATATANSALGEPGRDVLGLQESCLDGEWFS 120

DB 61 AWRNHTVPLRLRGREVPHSCRRYRLATATANSALGEPGRDVLGLQESCLDGEWFS 120

QY 121 QDYVLTSTVTEWNLVCEDDWKAPLTLSLFFVGLGSLFSTSGQLSDRFGKKNVLFVTMGQ 180

DB 121 QDYVLTSTVTEWNLVCEDDWKAPLTLSLFFVGLGSLFSTSGQLSDRFGKKNVLFVTMGQ 180

QY 181 TGFSEFQIFSKNFEMFVFLVFLVGMGOISNYAAAFVLGTEILGKSVRIIFSLGVCIFYA 240

DB 181 TGFSEFQIFSKNFEMFVFLVFLVGMGOISNYAAAFVLGTEILGKSVRIIFSLGVCIFYA 240

QY 241 FGYWVLPFLFAYFIRDWRMLIVALTMPGVLVCVALWFWIPESPRWLISQGFEEAEVIRKA 300

DB 241 FGYWVLPFLFAYFIRDWRMLIVALTMPGVLVCVALWFWIPESPRWLISQGFEEAEVIRKA 300

QY 301 AKANGIWPSTIFDPSELQDLSSKKQOOSHNIIDLRTWIRMTVIMSIMLMWTISVGYFG 360

DB 301 AKANGIWPSTIFDPSELQDLSSKKQOOSHNIIDLRTWIRMTVIMSIMLMWTISVGYFG 360

QY 361 LSLDTNHLGCDIFVNCFLSAMVEVPAYVLAWLLOQLPRYSKATALFLGGSVLLFMQLV 420

DB 361 LSLDTNHLGCDIFVNCFLSAMVEVPAYVLAWLLOQLPRYSKATALFLGGSVLLFMQLV 420

QY 421 PPDLYLATVLMVGMKFGVTAAFSMVYVYTAELYPTVVRNMGVSVSTASRLGSLISPYF 480

DB 421 PPDLYLATVLMVGMKFGVTAAFSMVYVYTAELYPTVVRNMGVSVSTASRLGSLISPYF 480

QY 481 VYLGAYDRFLPYILMGSLTILTAILTFLFPESFGTLPDITDMLRVKGMKHKRTPSHTR 540



Db 481 VYLGADYRFLPYLMGSLILRAILLFLPESGTFPLPTIDQMLRVKMKRKPSTHR 540  
 QY 541 MLKDQGERPTILKSTAF 557  
 Db 541 MLKDQGERPTILKSTAF 557

RESULT 2  
 OCN2\_MOUSE  
 ID OCN2\_MOUSE STANDARD; PRT; 557 AA.  
 AC Q92088;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,  
 DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).  
 GN SLC22A5 OR OCTN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=99113835; PubMed=9916797;  
 RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,  
 RA Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuishi T.,  
 RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,  
 RA Tsuji A.;  
 RT "Primary systemic carnitine deficiency is caused by mutations in a  
 RT gene encoding sodium ion-dependent carnitine transporter.";  
 RL Nat. Genet. 21:91-94(1999).  
 RN [2]  
 RC SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.  
 RC STRAIN=C3H;  
 RX MEDLINE=99057546; PubMed=9837751;  
 RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;  
 RT "A missense mutation of mouse OCTN2, a sodium-dependent carnitine  
 RT cotransporter, in the juvenile visceral steatosis mouse.";  
 RL Biochem. Biophys. Res. Commun. 252:590-594(1998).  
 RN [3]  
 RC SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC MEDLINE=99384224; PubMed=10454528;  
 RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,  
 RA Chen J., Conway S.J., Ganapathy V.;  
 RT "Functional characteristics and tissue distribution pattern of organic  
 RT cation transporter 2 (OCTN2), an organic cation/carnitine  
 RT transporter.";  
 RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
 CC -|- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE  
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
 CC CARNITINE.  
 CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -|- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL  
 CC STEATOSIS (JVS).  
 CC -|- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
 CC CATION SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

EMBL; AB015800; BAA36590.1; -;  
 EMBL; AF111425; AAC99787.1; -;  
 EMBL; AF110417; AAD54060.1; -;  
 MGI; MGI:1329012; SLC22a5.  
 InterPro; IPR003662; sub\_transporter.

Pfam: PF00083; sugar\_tr: 1.  
 PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1. Disease mutation.  
 KW Transport; Transmembrane; Glycoprotein; POTENTIAL.  
 FT TRANSMEM 21 41  
 FT TRANSMEM 143 163  
 FT TRANSMEM 173 193  
 FT TRANSMEM 198 218  
 FT TRANSMEM 233 253  
 FT TRANSMEM 258 278  
 FT TRANSMEM 342 362  
 FT TRANSMEM 374 394  
 FT TRANSMEM 407 427  
 FT TRANSMEM 431 451  
 FT TRANSMEM 489 509  
 FT CARBOHYD 57 57  
 FT CARBOHYD 64 64  
 FT CARBOHYD 91 91  
 FT CARBOHYD 322 322  
 FT VARIANT 352 352  
 SQ SEQUENCE 557 AA; 62779 MW; 6093F0EE9612B204 CRC64;

Query Match 87.2%; Score 2513; DB 1; Length 557;  
 Best Local Similarity 85.5%; Pred. No. 1.1e-154;  
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFTGEGWPPFORLFIFFLLSASIIPINGFTGLSSVFLIATPEHRCRVPDAANLSS 60  
 Db 1 MRDYDEVTAFTGEGWPPFORLFIFFLLSASIIPINGFTGLSSVFLIATPEHRCRVPDAANLSS 60  
 QY 61 AWRNHTVPLRLRDGREGVPHSCRRYRLATIANFSALEGPGRDVLQLEQESCLDGEWFS 120  
 Db 61 AWRNHSIPLKTKDGRQVQKCRRYRLATIANFSELGLEPGRDVLQLEQESCLDGEWFS 120  
 QY 121 QDYVLTSTVTEWNLCEDDWKAPLITISLFFVGLVLSFISQSLSDRFRGNKLVFTMGQ 180  
 Db 121 KDVFSLSTVTEWDLCKDDWKAPLITISLFFVGLVLSFISQSLSDRFRGNKLVFTMGQ 180  
 QY 181 TGFSELOFSKNFEMFVFLVGLVGQISNYVAARVLTGLGKSVRIIFSTGLGVCIFYA 240  
 Db 181 TGFSELOFSKNFEMFVFLVGLVGQISNYVAARVLTGLGKSVRIIFSTGLGVCIFYA 240  
 QY 241 FGMYVLPPLFAVEIRDRMLLVALTMPGVLVLCVLAWFIPESPRWLISQGRFEEAEVIRKA 300  
 Db 241 FGMYVLPPLFAVEIRDRMLLVALTMPGVLVLCVLAWFIPESPRWLISQGRFEEAEVIRKA 300  
 QY 301 AKANGIVVPSTIFDSELDSSKKQSHNLDLRTWNIRMTIMSLMWTISVGYFG 360  
 Db 301 AKANGIVVPSTIFDSELDSSKKQSHNLDLRTWNIRMTIMSLMWTISVGYFG 360  
 QY 361 LSLDTPNLHGDIYVNCFLSAMVEYVAVLWLLQYLPRLYSMAATLFGSVLLFMQLV 420  
 Db 361 LSLDTPNLHGDIYVNCFLSAMVEYVAVLWLLQYLPRLYSMAATLFGSVLLFMQLV 420  
 QY 421 PDLXYLATVLYMVGKFGVTAFAFMSVYVYTAELYPTVVRNMGVSVSSASRLGSLSPYF 480  
 Db 421 PSELYFLSTALVMVGKFGITSAYSVMYVYTAELYPTVVRNMGVSVSSASRLGSLSPYF 480  
 QY 481 VYLGADYRFLPYLMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKMKRKPSTHR 540  
 Db 481 VYLGADYRFLPYLMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKMKRKPSTHR 540  
 QY 541 MLKDQGERPTILKSTAF 557  
 Db 541 MLKDQGERPTILKSTAF 557

RESULT 3  
 OCN2\_RAT  
 ID OCN2\_RAT STANDARD; PRT; 557 AA.  
 AC Q70594; Q90WLO;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)

20-AUG-2001 (Rel. 40, Last annotation update)  
 ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,  
 MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER)  
 (UST2R) (CT1).  
 SLC22A5 OR OCTN2.  
 Rattus norvegicus (Rat).  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 (1)  
 SEQUENCE FROM N.A.  
 TISSUE=Kidney.  
 MEDLINE=98200080; PubMed=9541011;  
 Schoemig E., Spitzberger F., Engelhardt M., Martel F., Oerding N.,  
 Gruendemann D.;  
 "Molecular cloning and characterization of two novel transport  
 proteins from rat kidney.";  
 FEBS Lett. 425:79-86(1998).  
 (2)  
 SEQUENCE FROM N.A.  
 STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;  
 MEDLINE=99011422; PubMed=9792817;  
 Sekine T., Kusahara H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,  
 Kanai Y., Endou H.;  
 "Molecular cloning and characterization of high-affinity carnitine  
 transporter from rat intestine.";  
 Biochem. Biophys. Res. Commun. 251:586-591(1998).  
 (3)  
 SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 MEDLINE=99384224; PubMed=10454528;  
 Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,  
 Chen J., Conway S.J., Ganapathy V.;  
 "Functional characteristics and tissue distribution pattern of organic  
 cation transporter 2 (OCTN2), an organic cation/carnitine  
 transporter.";  
 J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
 (4)  
 FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE  
 TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
 INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
 CARNITINE.  
 (5)  
 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 (6)  
 TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES  
 AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND  
 ARTERIOLES IN THE HEART, IN THE LABYRINTHINE LAYER OF THE  
 PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE  
 BRAIN.  
 (7)  
 SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
 CATION SUBFAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 EMBL; AJ001933; CAA05106.1;  
 EMBL; AB017260; BAA34399.1;  
 EMBL; AF110416; AAD54059.1;  
 InterPro: IPR003662 sub\_transport.  
 Pfam: PF00083; sugar\_tr; 1.  
 PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
 Transport; Transmembrane; Glycoprotein.  
 TRANSMEM 21 41  
 TRANSMEM 143 163 POTENTIAL.  
 TRANSMEM 173 193 POTENTIAL.  
 TRANSMEM 198 218 POTENTIAL.  
 TRANSMEM 233 253 POTENTIAL.  
 TRANSMEM 258 278 POTENTIAL.  
 TRANSMEM 342 362 POTENTIAL.  
 TRANSMEM 374 394 POTENTIAL.  
 TRANSMEM 407 427 POTENTIAL.

FT TRANSMEM 431 451 POTENTIAL.  
 FT TRANSMEM 489 509 POTENTIAL.  
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 275 275 W -> G (IN REF. 2).  
 SQ SEQUENCE 557 AA; 62567 MW; 172472E7B0A5F043 CRC64;  
 Query Match 87.0%; Score 2509; DB 1; Length 557;  
 Best Local Similarity 85.5%; Pred. No. 2e-154;  
 Matches 476; Conservative 42; Mismatches 39; Indels 0; Gaps 0;  
 QY 1 MRDYDEVTAFLGEGWPFQRLIFLLSASIPNGTGLSSVFLIATPEHRCRVDPDAANLSS 60  
 DB 1 MRDYDEVTAFLGEGWPFQRLIFLLSASIPNGFNMSIVFLAGTPEHRCRVDPDAANLSS 60  
 QY 61 AWRNHTVPLRLDRGPHSCRRYRLATIANFSALEGPGRDVLQGLQESCLDGWEFS 120  
 DB 61 AWRNHSIPLETKDGROVPQSCRRYRLATIANFSALEGPGRDVLQGLQESCLDGWEYN 120  
 QY 121 QDVLSTIVTWNLCEDDKWAPLITSLFVGVLLGSFISQGLSDRFGKKNVLFVTGMQ 180  
 DB 121 KDVELSTIVTWNLCEDDKWAPLITSLFVGVLLGSFISQGLSDRFGKKNVLFVTGMQ 180  
 QY 181 TGSFLOIFSKNPFEMFVFLVGMGOISNYVAFLVGTILGKSVRIIFSTLGVCIYA 240  
 DB 181 TGSFLOIFSVNFEMFVFLVGMGOISNYVAFLVGTILGKSVRIIFSTLGVCIYA 240  
 QY 241 FGVNVLPLFAFIRDMRLMLVALTMPGVLCVLAWFIPSPRWLISQGRFEEAEVIRKA 300  
 DB 241 FGVNVLPLFAFIRDMRLMLVALTMPGVLCVLAWFIPSPRWLISQGRFEEAEVIRKA 300  
 QY 301 AKANGIVVPSTIFDPSLQDLSSKKQSHNILLDTNIRWVIMSLWMTISVGYFG 360  
 DB 301 AKANGIVVPSTIFDPSLQDLSSKKQSHNILLDTNIRWVIMSLWMTISVGYFG 360  
 QY 361 LSLDTPNLHGDIFVNCFLSAMVVPAYVLAWLLOVLPVRYSMATLFLGGSVLLFPMQLV 420  
 DB 361 LSLDTPNLHGDIFVNCFLSAMVVPAYVLAWLLOVLPVRYSMATLFLGGSVLLFPMQLV 420  
 QY 421 PDLIYLAIVLVVGVKFGVTAFAFVSMVYVTAELVPTVVRNMGVGVSTASRLGSLSPYF 480  
 DB 421 PDLIYLAIVLVVGVKFGVTAFAFVSMVYVTAELVPTVVRNMGVGVSTASRLGSLSPYF 480  
 QY 481 VYLGAYDRELPIYLMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKRTPSHFR 540  
 DB 481 VYLGAYDRELPIYLMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKRTPSHFR 540  
 QY 541 MLKDGQGRPTILKSTAF 557  
 DB 541 TQKDGESPTVLKSTAF 557  
 RESULT 4  
 YLX5\_CAEBL STANDARD; PRT: 751 AA.  
 AC P46501;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE HYPOTHETICAL 84.8 KDA PROTEIN F23F12.5 IN CHROMOSOME III.  
 GN F23F12.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2.  
 RA Du Z.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U12965; AAA20607.1; -;  
DR WormPep; F23F12.5; CE01252;  
DR InterPro; IPR002184; Srib.  
DR InterPro; IPR003662; sub\_trnsporttr.  
DR Pfam; PF02175; Srib; 1.  
DR Pfam; PF00083; sugar\_tr; 1.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 24 44 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT TRANSMEM 102 122 POTENTIAL.  
FT TRANSMEM 137 157 POTENTIAL.  
FT TRANSMEM 200 220 POTENTIAL.  
FT TRANSMEM 235 255 POTENTIAL.  
FT TRANSMEM 320 340 POTENTIAL.  
FT TRANSMEM 348 368 POTENTIAL.  
FT TRANSMEM 379 399 POTENTIAL.  
FT TRANSMEM 410 430 POTENTIAL.  
FT TRANSMEM 432 452 POTENTIAL.  
FT TRANSMEM 515 535 POTENTIAL.  
FT TRANSMEM 547 567 POTENTIAL.  
FT TRANSMEM 583 603 POTENTIAL.  
FT TRANSMEM 614 634 POTENTIAL.  
FT TRANSMEM 678 698 POTENTIAL.  
SQ SEQUENCE 751 AA; 84632 MW; A5C4F43540295EFC CRC64;

Query Match 13.4%; Score 387.5; DB 1; Length 751;  
Best Local Similarity 25.8%; Pred. No. 6.4e-18;  
Matches 106; Conservative 89; Mismatches 189; Indels 27; Gaps 9;

QY 127 TIVTEWNLVC-EDDWKAPLTISLFVGVLLGSFISGQSDRFGKRVLFVTMGQTGFSF 185  
Db 303 SSVQDFRMFCCTKAYDAAWAVATQIFGILGAIYGHGDFGRKPVFFGISVGLFGV 362  
QY 186 LQIFSKNEFMFVFLVGVGMQISNYAAVFLGVTEILGKSVRIIFSILGVGICFVAFGY-- 243  
Db 363 ASGFAPSEVFAAFVIGTSTIASILIVFYAYILEFTEPEQRPVLR-----FFNWGIAR 417  
QY 244 MVLPLFAFYFIRDRMLLVALTMPGVCVALWMFIPESPRWLSQGRPEEAPEVILIRKAAKA 303  
Db 418 LVFTLACFCYGRSAAIATSLSLPLPVLILPESPKWNTKKRFRDARAEEKRVAWL 477  
QY 304 NGIVVPSTIFDPELOD-LSSKKQOOSHNLIDLLRTWIRMTVIMSLMWLTISVGYFGLS 362  
Db 478 SGIPVYNDQDSIESEKLEKSKYIKTMKDLFTSWTIAYRTIIVGVLSFSTLSAFGSD 537  
QY 363 LDTPNLHGDIFVNCFLSAMVVPAYVAVLAWLLQVLP-----RRYSM--ATAFLGGSVLLF 416  
Db 538 LNSGNAGNFVLSQVSGAVTAFKIFVFLDVTVPFDRRLRHOYQIAMLICYCVIMV 597  
QY 417 MOLVPP-----DLYLATVLMVGKFGVTAFAFSVMVYVYTAELYPTVVRNMGVGVSS 467  
Db 598 LMLPESDCGSGSDRLAI--IINIIGVSFIETWDACYLVAVCEPPTKIRTIGIGTCS 655  
QY 468 TASRLGSILSPFYVLGAYDRFLPYLMGSL-TILTALLTLFLPESGTP 517  
Db 656 LIARTGALLQPMAYLSDIYPAPYAVVCISGIGTISLLISCFLPDTRKGVDL 706

RESULT 5  
YMP3\_CAEEL STANDARD; PRT; 1222 AA.  
AC Q10947;  
DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHEICAL I39.9 KDA PROTEIN B0361.3 IN CHROMOSOME III.  
GN B0361.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RL Du Z.;  
RA Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U00031; AAK18864.1; -;  
DR WormPep; B0361.3; CE00752.  
DR Pfam; PF00083; sugar\_tr; 1.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 51 71 POTENTIAL.  
FT TRANSMEM 148 168 POTENTIAL.  
FT TRANSMEM 182 202 POTENTIAL.  
FT TRANSMEM 232 252 POTENTIAL.  
FT TRANSMEM 263 283 POTENTIAL.  
FT TRANSMEM 307 327 POTENTIAL.  
FT TRANSMEM 329 349 POTENTIAL.  
FT TRANSMEM 355 375 POTENTIAL.  
FT TRANSMEM 382 402 POTENTIAL.  
FT TRANSMEM 413 433 POTENTIAL.  
FT TRANSMEM 442 462 POTENTIAL.  
SQ SEQUENCE 1222 AA; 139868 MW; CBA4ZA80A254FB80 CRC64;

Query Match 12.4%; Score 357; DB 1; Length 1222;  
Best Local Similarity 21.1%; Pred. No. 9.6e-16;  
Matches 118; Conservative 106; Mismatches 209; Indels 126; Gaps 14;

QY 3 DYDEVTAFIGEWGPPORLIFFLSASIIIPNGFTGLSSVFLIA---TPHRCRVDPDAWL 58  
Db 35 DPDKFVEAYGAYGKYQIFTYVLVQTL---NFFYSSSMYIMSFVQLNLEKQCE----- 83  
QY 59 SSAMRNHTVPLRLDRDREVPKSCRRYRLATIANFSALGLEPGRDVLGQLEQESCLDGWE 118  
Db 84 ---YKNETIP-----ISETCQ-----IETESSKAFGNLNGEYC-----G 114  
QY 119 FSDQVVL-----STIVTEWNLVGEDDWKAPLTISLFVGVLLGSFISGQSDRFGK 170  
Db 115 IAETLVNVTNQKASINLLVDFDLSCHWFQFGLTIFTIGAVIAPFMSMLADRYGRK 174  
QY 171 NVLFVTMGQTGFSFTQIFSKNFEMFVFLVGVGMQISNYAAVFLGVTEILGKSVRIIF 230  
Db 175 PIIVTTAILAFLANMAASFNPFAIFLILRAFTGACSDSVLSVASVATCEYLSEKAR-AW 233  
QY 231 STLGVCFIYAFGYVWLPLFAFYFIRDRMLLVALTMPGVCVALWMFIPESPRWLSQGRF 290  
Db 234 ITVVYNVMSLGMWTTLLVLTMTDDWRWRYFIVSLPGVYGFALWYFLPSPHWTITKNT 293  
QY 291 EEAETVIRKAAKANGIWPSTIFDPSELQDLSKKQOOSHNLIDLLRTWIRMTVIMSIL 350  
Db 294 EKLKVKYIKTANR----- 305  
QY 351 WMTISVGYFGLSLDTPNLHGDIFVNCFL-SAMVEVPAYVAVLAWLLQYLPVRYSM----- 403  
Db 306 -WYISLVYFAISPMSELGSD-QVQAFLYSSLLIEIPAGLAVIPMMKMGKRMKVIWCLVF 363



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	
EMBL; U29579;	AAA69281.1; ALT_INIT.
DR EMBL; AE000360;	AAC75813.1; ALT_INIT.
DR EcoGene; EG13126;	YGCs.
DR InterPro; IPR003662;	sub_Lrnspstr.
DR Pfam; PF00083;	sugar_tr_1.
DR PROSITE; PS00216;	SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217;	SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein;	Transport; Inner membrane;
KW Complete proteome.	
FT TRANSMEM	23 43 POTENTIAL.
FT TRANSMEM	57 77 POTENTIAL.
FT TRANSMEM	86 106 POTENTIAL.
FT TRANSMEM	115 135 POTENTIAL.
FT TRANSMEM	143 163 POTENTIAL.
FT TRANSMEM	176 196 POTENTIAL.
FT TRANSMEM	254 274 POTENTIAL.
FT TRANSMEM	287 307 POTENTIAL.
FT TRANSMEM	312 332 POTENTIAL.
FT TRANSMEM	338 358 POTENTIAL.
FT TRANSMEM	370 390 POTENTIAL.
FT TRANSMEM	401 421 POTENTIAL.
SQ SEQUENCE	445 AA; 48234 MW; B59B452721B15774 CRC64;

Query Match 10.7%; Score 308; DB 1; Length 445;  
Best Local Similarity 26.6%; Pred. No. 4.8e-13;  
Matches 102; Conservative 79; Mismatches 168; Indels 34; Gaps 11;

QY	151	VGVLGSFISQGLSDRGKXNVLFTVMGMQGTGFSFLOIFSKNFEMFVFLVLMGMQISN	210
Db	65	LGLFLSVLGSWISDIXGROKIFETFSFLITLASLFOFFATTPEHLIGRLIGLIGLGD	124
QY	211	YVAAFVLGTEILGSKSVRIISTLGV-CIEYAFGYMWLPLFA-YFIRD- ---WRMLVALT	264
Db	125	YSVGHTLLAEFSRPHRGIL- --LGAFSVVMVTGVYLIASIAGHHFISENPEAKRWLLASAA	182
QY	265	MPGVLCVALWMFIPESPRWLISOGREEAEVIRKAAKANGIVVPSTIFDPSEL- --QDLS	322
Db	183	LPALLITLLRWGTPESPRWLLROGFAEAAHVHR- ---YFSPHVLGDEVV	230
QY	323	SKQCSHNIIDLRTNIRNMTVMSIMLWMTISVGVFGSLDTPNLHGDIIVNCFLSAMV	382
Db	231	TATHKHIKTIFSSRYWR- --RTAFNSVFVCLVPIPWFEVITVTLPTAQITIGLEDALTASL	287
QY	383	EVPAAYLAWLLQYLPFRYSMATALFSGSVLFL- ---MQLVPPDLYIATVLMVGVKF	437
Db	288	MLNALLTVGALLG- LVLTHLLAHRKFLGSLFLLAALTVMMACLPSGSSLTLLFLVLFST	346
QY	438	GVTAAFSVMVYVTAELYPVTVRNMNGVGSASTRGLSGILSPAFV- --YLGAYDRFLPYILM	495
Db	347	TISAVSNLGVLPAESFPTDIRSLGVGFATAMSRIGAAVSTGLLPVWLQWGMQVTLILL	406
QY	496	GSUTILTALT-LFLPESFGTPL	517
Db	407	ATVLLGVFVVYTWLWAPETRALPL	429

RESULT	8
YTI3_CAEEL	
ID	YTI3_CAEEL
STANDARD;	
PRT;	435 AA.
AC	Q10917;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	01-OCT-1996 (Rel. 34, Last annotation update)
DE	HYPOPHETICAL 48.6 KDA PROTEIN B0252.3 IN CHROMOSOME II.
GN	B0252.3.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; .
OC	Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=BRISTOL N2;  
 RA Du Z., Waterston R.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement>).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC

CC	EMBL; U23453; AAC46757.1; -
DR	WormPep; B0252.3; CE02419.
DR	InterPro; IPR03662; sub_trnspospr.
DR	Pfam; PF00083; sugar_tr; 1.
KW	Hypothetical protein; Transmembrane.
KW	Hypothetical protein; Transmembrane.
FT	TRANSMEM 55 75 POTENTIAL.
FT	TRANSMEM 384 404 POTENTIAL.
SQ	SEQUENCE 435 AA; 48571 MW; 576160359FA1840B CRC64;

Query Match	10.1%	Score 291.5	DB 1	Length 435
Best Local Similarity	21.5%	pred. NO. 5.4e-12		
Matches 90	Conservative	78	Mismatches 168	Indels 83
				Gaps 8

Qy	128	I V T E N W L C V E D W K A P I T I S I F V G V I L L G S F I S Q L S D R F G R K N V L E V T M G M T G S F L Q
Db	91	V A D E F D I T G D A S W L A E S T T F Y M V G N I G G M F I P L A D H Y G R L P V F V A T V L L A V G G M I S
Qy	188	I F S K N F E M F V V L F V L V G M G Q I S N V A A F V L G T E I L G K S V R I I I S T L G C V C I F Y A F G Y M V L P
Db	151	A F S T I M M F C I M R M I H G I F T A A G L A G W I G Y E N T P U L R - F T S V Y F G Y M V W V G A C F L G
Qy	248	L F A T F I R D R M L L V A L T M P G V - L C V A L W M F I P E S P R W I I S O G R F E A E V I I R K A A R A N G I
Db	210	L L A Y I L P D W R Y L M F C I S V P N I F V A L L I Y M V P E S L H F L V S Q Q N E K T E A W L E K I R G P K G D
Qy	307	V V P T I E D P S E L Q D L S K K Q O S H N I L D L R T W N I R M Y T I M S I M L W M T I S V G Y F G L S L D T P
Db	270	I S A S D I -----V E D R E N G S S F K T L C R E-----
Qy	367	N L H G D I F V N C F L S A M V E P A V V - L A W I L L Q Y L P R Y S M A T A L F L G S V I L F M Q L V P D P L Y
Db	293	-----I K T F K T L F O D R R Y I L F W I-----G I L L Y E F G R K P L F E
Qy	426	Y L A T V L W M V G K G Y T A A F S M V Y V V T A E L Y P T V R N M G V G S V S T A S R L G S I L S I F V Y V L G A
Db	327	F C A -----H E F G R S S L H F S D F H E Q I F P T D G E N K C I G E C E T L S R F G G M L S P Y L S H L T A
Qy	486	Y D R E L P Y I L M G S L I L F A I L T F L P E S F G T P L P D T I D O M L K V K G M K H R K T P S T H R M L K D
Db	381	V H A L A P A I T S L I A V S G G L L T I L P E T I N T K L P S T I-----A E T A S R L O I D D

RESULT	9
YCEI_BACSU	
ID_YCEI_BACSU	STANDARD; PRT; 400 AA.
AC	O34691;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YCEI.
GN	YCEI.
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=1423;
RN	[1]

```
RP SEQUENCE FROM N.A.
RC STRAIN=166;
RA Kumano M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB000617; BAA22256.1; -
CC DR EMBL; 299105; CAB12089.1; -
CC DR Subtilist; BG12773; ycel.
CC DR InterPro; IPR003662; sub.transprtr.
CC DR Pfam; PF00083; sugar.tr.1.
CC DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
CC DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
CC KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
CC FT TRANSMEM 15 35 POTENTIAL.
CC FT TRANSMEM 50 70 POTENTIAL.
CC FT TRANSMEM 78 98 POTENTIAL.
CC FT TRANSMEM 99 119 POTENTIAL.
CC FT TRANSMEM 143 163 POTENTIAL.
CC FT TRANSMEM 166 186 POTENTIAL.
CC FT TRANSMEM 218 238 POTENTIAL.
CC FT TRANSMEM 254 274 POTENTIAL.
CC FT TRANSMEM 281 301 POTENTIAL.
CC FT TRANSMEM 305 325 POTENTIAL.
CC FT TRANSMEM 344 364 POTENTIAL.
CC FT TRANSMEM 371 391 POTENTIAL.
CC SQ SEQUENCE 400 AA; 43708 MW; EOAE0CEB5DD27395 CRC64;

Query Match 10.0%; Score 287; DB 1; Length 400;
Best Local Similarity 24.9%; Pred. No. 9.7e-12;
Matches 104; Conservative 76; Mismatches 161; Indels 76; Gaps 18;

QY 116 GNEF-SQDV-YLSTIV----TEWNLVCD-DWKAPLTISLFVGVLLGSFTSGOLSDRF 168
DB 19 GWFDAVDGILSFIAALHVENNLSPPEMKWG----SVNSIGMAGAFLEGLADRI 74
QY 169 RKNVLFVTMGQTGFSFLQIFSKNFEMFVLVGVGMQISNYVAAFVLGTEIL-----G 223
DB 75 RKKVFITILLCFSIGSGISAFVTSLSAFLILFVIGMGLGELPVASTLVSEAVVPEKR 134
QY 224 KSVRIEFTLGVCIYAFGYVWLPLEAFVI---RDWRMLLVALLTMPGVLCVALWFFPES 280
DB 135 RVIVLES-----FWVGWLAALISVIFPFGWQAALLTLTAFYALYLRSLPDS 188
QY 281 PRLWISQGRFEAEVIRKAANGIVVPSTIFDPSELQDLSKKQOS--HNILDLLRTWN 339
DB 189 PKY-----ESLSAKKRWENVKSVWARY 213
QY 340 IRNVTIMSLMWTISVGYFGLSLDTPN---LHGDIFVNCF----LSAMVEVPYVLAWL 392
DB 214 IRPTVMSI-VNFCVVFYSYGNLWLPSPVMLLKGFMSIQSFYVLLMTLAOLPGYFSAW 272
QY 393 LQQLPRYSNATALF-LGGSVLLFMQLVPPDLVYLVVGMVKGFTAFSVVYVTA 451
DB 273 LIEKAGRWILVYVLIAGTSAFYFG--TADSLSLLLTAGVLLSFFNL-GAWGVLYATP 329
QY 452 ELYPTVVRNMGVSGSTASRLSGISLSPYFV-YLGAYDRFLPYILMGSITLITAILTL 507
DB 330 EQYPTPAIRATSGTFAAFGRIGGIFGLLVGTAA--RHISFSVIFSFCITAILAV 384
```

```
RESULT 10
YAAU_ECOLI STANDARD; PRT; 443 AA.
AC P31579; P31578; P75628;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YAAU.
GN YAAU OR B0045.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
CC IN POSITION 142 AND 232 THAT PRODUCE TWO SEPARATE ORFS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D10483; BAA01321.1; ALT_FRAME.
CC DR EMBL; AE000114; AAC73156.1; -
CC DR PIR; S40566; S40566.
CC DR EcoGene; EGI1566; yaaU.
CC DR InterPro; IPR003662; sub.transprtr.
CC DR Pfam; PF00083; sugar.tr.1.
CC DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
CC DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
CC KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
CC Complete proteome.
CC FT TRANSMEM 19 39 POTENTIAL.
CC FT TRANSMEM 54 74 POTENTIAL.
CC FT TRANSMEM 85 105 POTENTIAL.
CC FT TRANSMEM 114 134 POTENTIAL.
CC FT TRANSMEM 146 166 POTENTIAL.
CC FT TRANSMEM 174 194 POTENTIAL.
CC FT TRANSMEM 242 262 POTENTIAL.
CC FT TRANSMEM 283 303 POTENTIAL.
CC FT TRANSMEM 323 343 POTENTIAL.
CC FT TRANSMEM 374 394 POTENTIAL.
CC FT TRANSMEM 402 422 POTENTIAL.
CC FT CONFLICT 7 7 F -> L (IN REF. 1).
CC FT CONFLICT 12 16 FSSIH -> CLHY (IN REF. 1).
CC FT CONFLICT 30 30 D -> Y (IN REF. 1).
CC FT CONFLICT 115 115 L -> F (IN REF. 1).
CC FT CONFLICT 132 132 T -> N (IN REF. 1).
```

```

FT CONFLICT 240 240 R -> C (IN REF. 1).
FT CONFLICT 304 304 N -> S (IN REF. 1).
FT CONFLICT 395 395 N -> D (IN REF. 1).
SQ SEQUENCE 443 AA; 48666 MW; AEF484D8B109DE05 CRC64;

Query Match 9.3%; Score 267; DB 1; Length 443;
Best Local Similarity 24.6%; Pred. No. 2.1e-10;
Matches 104; Conservative 81; Mismatches 184; Indels 54; Gaps 13;

QY 134 LVCEDDWKAFLTISLFVGVLLGSFTS-----GQSDRFGRKKNVFLVTMGMTGFSFLQIF 189
Db 48 LKLDADM-----IGLLGAGTLAGLVGTSLFQISDKVGRKRMFLDIIAIGVISVATMF 102
QY 190 SKNFEMFVFLVGVGOISNYAAVFLGTEILGKSVRIIFSTGLVCIFYAFGYVWLPLF 249
Db 103 VSSPVELLMVRVLIGVIGADYPIATSMITFSSTRQ-AFSISFIAAMWYVGATCADLV 161
QY 250 AYFRD---WRMLVALTMPGVLCVALWFIPESPRWLISQGRPEEAETIRKAAKANG 305
Db 162 GYWDYVEGGWRMGLGSAIPCLLIILGREFELPESRWLLRGRVKECEEMIK----- 215
QY 306 IVPSTIFDPSELQ-----DLSSKKQOQSHNLDLRTWIRVMTIMSIMLWMTISVGYFG 360
Db 216 LFCEPVAFDEEQQOTRFRDLNRRHPPF-VLFVAAIWTQVTPMFAIYTFGQIVGLLG 274
QY 361 LSLDTPNLHGDIKFNCFLSAMVEPAVLANLLQLVLPKRYSWATA-----LFLGGSVLLF 416
Db 275 LGVKNAALGNVVISLF-----FMLGCIPTPMLNLTAGRRPLLGSPFAMMT 320
QY 417 MOLVPPDLYLAVLVVMGVKFGVTAIFS-----MVVYVYTAELPTVVRNMGVSVSTASR 471
Db 321 LAVALGLIPDMGLWLVMAFAVAFSFGGPNLQWLINELPPTDIRASAVIMSLSR 380
QY 472 LGSILSPYF--VVLGAYDRELPYLMGSLTILTAIL--TLFPESEGTPLPDTIDQMLRV 527
Db 381 IGTIVTWPALPIFINNYG-ISNTMLMGAGISLFGLLISVAFAPETRGMSLAQTSNMTIRG 439
QY 528 KGM 530
Db 440 QRM 442

RESULT 11
YDJK_ECOLI STANDARD; PRT; 459 AA.
AC P76230; P76911;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YDJK.
GN YDJK OR B1775.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,

```

```

RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA Sequence of the Escherichia coli K-12 Genome
DNA Res. 3:363-377(1996).
CC Corresponding to the 28.0-40.1 min Region on the Linkage Map.";
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE000272; AAC74845.1; .
DR EMBL; D90821; BAA15573.1; ALT_INIT.
DR EcoGene; EGI3487; ydjk.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 154 174 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT TRANSMEM 421 441 POTENTIAL.
SQ SEQUENCE 459 AA; 49602 MW; BC8AB53ECB8DD77 CRC64;

Query Match 9.2%; Score 265; DB 1; Length 459;
Best Local Similarity 26.2%; Pred. No. 2.9e-10;
Matches 101; Conservative 79; Mismatches 153; Indels 52; Gaps 18;

QY 152 GVLGSGTSGQLSDRFGRKKNVFLVTMGMTGFSFLQIFSKNFEMFVFLVGVGOISNY 211
Db 70 GMFLGALVGGIIGDKTRGNFALYEAIHIASMVVGVAFSPNMDFLIACRFVNGVGLGALL 129
QY 212 VAAFLVGLTEIL-GK-----SVRIIFS---TLGVCIFYAFGYVWLPLFAFYIRDMRLVA 262
Db 130 VILFAGTYMFGNRGTWSSRVFIGNWSYPLCSLIANG--LTP LIS-AEWNVRVQLLI 186
QY 263 LTPMGVLCVAL-WFPIESPRLWISQGRPEEAETIRK-----AAKANGIVVPSTIFDSE 317
Db 187 PALSLSIATLAWRYFPSPRWLESRGYQAEKVMRSIEEGVIRQTGKPLP-----PW 241
QY 318 LODLSKKQOQSHNLDLRTWIRVMTIMSIMLWMTISVGYFGLSLDTP-----NLH 369
Db 242 IAD-DGKAPQAPVPSALLTVGLLKRVLGSCVL-IAMNVVQVTLINWLPTIFMTGGINLK 299
QY 370 GDFVNCFLSAMVEVP-AVLAWLLOLPLRPYSMATALLFLGSGVLLFMQLVPPDLYYLA 428
Db 300 DSIVLNTM--SMFGAPFGFIAMLVMDKIPRK-TMGVGLLLIILAVLG-----IYSLQ 349
QY 429 T---VLVMVGKFGVTAAFSNV-----YVYTAELPTVVRNMGVSVSTASRLGISLSPY-- 479
Db 350 TSMLLITLIGFFLITFVYVYVVCASAVYVPEIWPTEAKLRGSGLANAVGRISGIAAPYAV 409
QY 480 FVYLGAIDRELPYLMGSLTILTAI 504
Db 410 AVLLSSYGVGTGVFILLGAVSIIVAI 434

```



RESULT 12

SVY2\_RAT STANDARD; PRT; 742 AA.  
 ID Q02563;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE SYNAPTIC VESICLE PROTEIN 2 (SV2).  
 GN SV2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.  
 RC TISSUE=Brain;  
 RX MEDLINE=92390722; PubMed=1519064;  
 RA Bajjalieh S.M., Peterson K., Shingal R., Scheller R.H.;  
 RT "SV2, a brain synaptic vesicle protein homologous to bacterial  
 transporters.";  
 RL Science 257:1271-1273(1992).  
 CC -1- FUNCTION: MAY ACT AS A VESICLE-LOCALIZED NEUROTRANSMITTER  
 CC TRANSPORTER OR MAY FUNCTION AS AN ION TRANSPORTER OR CHANNEL.  
 CC -1- SUBCELLULAR LOCATION: SYNAPTIC VESICLE.  
 CC -1- TISSUE SPECIFICITY: NEURAL AND ENDOCRINE CELLS OF BRAIN AND  
 CC SPINAL CORD.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; L05435; AAA42188.1; -;  
 DR Pfam; PF00083; sugat\_tr; 1;  
 KW Synapse; Nerve; Glycoprotein; Neurotransmitter transport;  
 KW Transmembrane.  
 FT DOMAIN 1 163  
 FT TRANSMEM 164  
 FT DOMAIN 183  
 FT TRANSMEM 205  
 FT DOMAIN 226  
 FT TRANSMEM 243  
 FT DOMAIN 261  
 FT TRANSMEM 283  
 FT DOMAIN 294  
 FT TRANSMEM 322  
 FT DOMAIN 334  
 FT TRANSMEM 335  
 FT DOMAIN 356  
 FT TRANSMEM 446  
 FT DOMAIN 470  
 FT TRANSMEM 593  
 FT DOMAIN 612  
 FT TRANSMEM 627  
 FT DOMAIN 648  
 FT TRANSMEM 650  
 FT DOMAIN 670  
 FT TRANSMEM 694  
 FT DOMAIN 712  
 FT TRANSMEM 713  
 FT DOMAIN 732  
 FT TRANSMEM 742  
 FT CARBOHYD 498  
 FT CARBOHYD 548  
 FT CARBOHYD 573  
 SQ SEQUENCE 742 AA; 182705 MW; 565DE7EF2929D5DB CRC64;

Query Match

9.18; Score 262.5; DB 1; Length 742;

Best Local Similarity 17.5%; Pred. No. 6.8e-10;  
 Matches 115; Conservative 103; Mismatches 173; Indels 267; Gaps 21;  
 QY 96 GLEPGRDVLDGQEQESCLDGEWFSQDYLSTIVTE-----WNL----- 134  
 DB 125 GLSDGEGPPGGRGEAQRKDRLELAQQ--YETILRECGHGFRQWTLFVLGLALMADGVE 182  
 QY 135 -----VCEDDWKAPLTISLFFVGLVGLSGISQSLSDRFRGNKLVFTVMGMQ 180  
 DB 183 FVVGCVLPSAEKDMCLSDSNKGMGLIVLGMVMGAPLWGLGLADRLGRROCLLISLVN 242  
 QY 181 TGFSLQLQFSKNFEMFVFLVFLVGMGOISNVYAAVFLGTEILGKSVRIEFTLIG----- 234  
 DB 243 SVFAFSSFSVGVGYTFPCRLLSVGV-----IGGSIPVIVFSYFSEFLAQE 287  
 QY 235 -----VCIFYAFG-----YMLPLF-----AYFIRDWRMLLVALTMPGVLC 270  
 DB 288 KRGEHLNWLCLMFMIGMIGVYAAAMAWAIIPHYGWSFQMGSAVQFHSWRVFLVFAFPVSFA 347  
 QY 271 VALWVFIPEPRWLISQCFEEAEVIRKA-----AKANGIVVPSTIFDPSLQDLSSKKQ 326  
 DB 348 IGAULTQPSRPFLENGKHDEAWMLKQVHDTHNRAG--HPERVFSVTHIKTI-----H 401  
 QY 327 QSHNILDL-----LRTWNI-----RMVTIMSIMLWMTISVGY 358  
 DB 402 QEDLIEIQSDTGTWYQWGVRAVLSLGGQWGNFLSCFSEPEYRITLMMMGVWFTMSFSY 461  
 QY 359 FGLSLDTPNL-----HGDIFVNC----- 376  
 DB 462 YGLTWFPDMIRHLQAVDYAARTKVFGEVHEVTNFLENQIHRGGQYFNDFKFIGLRL 521  
 QY 369 -----HGDIFVNC----- 376  
 DB 522 KSVSFEDSLFEECYFEDVTSSNTFRNCTFNTVYNTDLFEYKFNVSRLVNSTFLLHKE 581  
 QY 377 -----FLSAMVEVPAYVLAWLQLLVLPVRYNSWATALELGGSVL-- 414  
 DB 582 GCPLDVTGTGSGAYMVYFVSLFGLAVLPNGIVSALLMDKIGRLMLA-----GSSVLS 636  
 QY 415 ---LPMQLVPPDLYLATVLMVWKGFGVT-AAFSMVYVYTAELYPTVVRNMGVGVSTAS 470  
 DB 637 VSCFFLSFGNSESAMIA-LLCLEF--GVSIASNALDVLTVLPYPSDKRTTAFGFLNALC 693  
 QY 471 RLGSIL--SPYFVVLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPPTIDQMLR 526  
 DB 694 KLAIVLGISITSFVG-----ITKAAPIIFASAAALGSSLAKLPETRGQVLIQ 742  
 RESULT 13  
 YFIG\_BACSU STANDARD; PRT; 482 AA.  
 ID YFIG\_BACSU  
 AC P34723;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YFIG.  
 GN YFIG.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96262713; PubMed=8704981;  
 RA Yamamoto H., Uchiyama S., Fajar A.N., Ogasawara N., Sekiguchi J.;  
 RT "Determination of a 12 kb nucleotide sequence around the 76 degrees  
 RT region of the Bacillus subtilis chromosome.";  
 RL Microbiology 142:1417-1421(1996).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D50543; BAA09111.1; -  
 DR EMBL; Z99108; CAB12635.1; -  
 DR Subtilist; BG11854; yfiG.  
 DR InterPro; IPR003663; Sugar\_transport.  
 DR InterPro; IPR003662; sub.transports.  
 DR Pfam; PF00083; sugar.tr; 1.  
 DR PRINTS; PR00171; SUGTRNSPORT.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 30 50  
 FT TRANSMEM 60 80 POTENTIAL.  
 FT TRANSMEM 93 113 POTENTIAL.  
 FT TRANSMEM 121 141 POTENTIAL.  
 FT TRANSMEM 156 176 POTENTIAL.  
 FT TRANSMEM 185 205 POTENTIAL.  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT TRANSMEM 302 322 POTENTIAL.  
 FT TRANSMEM 332 352 POTENTIAL.  
 FT TRANSMEM 353 373 POTENTIAL.  
 FT TRANSMEM 401 421 POTENTIAL.  
 FT TRANSMEM 424 444 POTENTIAL.  
 FT SEQUENCE 482 AA; 52756 MW; 90851C4F4C48EE01 CRC64;

Query Match 8.9%; Score 256; DB 1; Length 482;  
 Best Local Similarity 23.9%; Pred. No. 1.2e-09;  
 Matches 105; Conservative 83; Mismatches 167; Indels 84; Gaps 19;  
 QY 144 ITISLFVGVLLGSFISQSLDRGRKN-----VLFFV--TMGMQGFSGFLQIFSKNFEM 195  
 DB 64 LVASSLLGAAGFAMFGGRGSRKRTLYLALLFIAATLGC-----TFSPNASV 115  
 QY 196 FVFLFVLGM--GQISNVAAFLVGLTEILCKSVRII-----FSTLGVCFYAFGYWVLP 248  
 DB 116 MIAFRFLGLAVGCASVTPTFLAETSPAERRGRIVTQNELMIVIGQLLAYTFNALIGST 175  
 QY 249 FAYFIRDMRLVLTALTPGVLCLVLMW---FIPSPRLWISQGRFEAEVIRKAANKANG 305  
 DB 176 MGESANVRWMLVATLPVAV---LWFGMLIVPESPRWLAAGRMGDALRVLRQIRE--- 229  
 QY 306 IVPSTIFDPSELQD-----SSKKQOOSHNIIDLRTNIRMTI---MSIMLWMT 353  
 DB 230 -----DSQAQOEIKHAIEGTAKKAGFH---DFQEPWIRRLIFIGIGIAIVQOIT 278  
 QY 354 --ISGVFGSL-----DTPNLHGDFVNCFLSAMVEVPAYVLAWLILLYLPRR--- 400  
 DB 279 GVNSIMTYGTEILREAGQFTEAALIGNI-----ANGVISVIAVIFGILWLGKVRRRPMLI 333  
 QY 401 ---YSMATPLGSLVLEWOLVPPDLYLATVLMVGKFGVTAAGSMVYVYTAELYPTV 457  
 DB 334 IQGIGTALLIGLSILECTPALYPVVLSTILFLAQQTATISVITWMLSEIFPMH 393  
 QY 458 VRNMGVGVSS-----TASRLGSLSPYFV-YLGAYDRFLPYLMGSLTILFAITLFLPES 512  
 DB 394 VRGLMGISTECLTANFLGTFEPILLNHIGMSATFFIEVAMNAIL--FVKVYVPT 451  
 QY 513 FGTPLPIDQMLRVKGMK 531  
 DB 452 KGRSL-EQLEHSFRQYGR 469  
 RESULT 14  
 YDJE\_ECOLI  
 ID YDJE\_ECOLI STANDARD; PRT; 452 AA.  
 AC P38055; P77244;

DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YDJE.  
 CC YDJE OR B1769.  
 CC Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CC NCBI\_TaxID=562;  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=K12 / MG1655;  
 RC MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=K12;  
 RC MEDLINE=97251357; PubMed=9097039;  
 RX Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitagawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 RN [3]  
 RP SEQUENCE OF 360-452 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=89357501; PubMed=2670682;  
 RA Jerlstrom P.G., Bezjak D.A., Jennings M.P., Beacham I.R.;  
 RT "Structure and expression in Escherichia coli K-12 of the  
 L-asparaginase I-encoding ansA gene and its flanking regions.";  
 RL Gene 78:37-46(1989).  
 RN [4]  
 RP IDENTIFICATION.  
 RP MEDLINE=95075659; PubMed=7984428;  
 RA Borodovsky M., Rudd K.E., Koonin E.V.;  
 RT "Intrinsic and extrinsic approaches for detecting genes in a  
 bacterial genome.";  
 RL Nucleic Acids Res. 22:4756-4767(1994).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE000272; AAC74839.1; -  
 CC EMBL; D90820; BAA15560.1; -  
 CC EMBL; D90821; BAA15567.1; -  
 CC EMBL; M26934; -; NOT\_ANNOTATED\_CDS.  
 CC EcoGene; EG12369; ydje.  
 CC InterPro; IPR003662; sub.transports.  
 CC Pfam; PF00083; sugar.tr; 1.  
 CC PROSITE; PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
 CC PROSITE; PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 KW Hypothetical protein; Transmembrane; Inner membrane;  
 KW Complete proteome.  
 FT TRANSMEM 23 43 POTENTIAL.  
 FT TRANSMEM 60 80 POTENTIAL.  
 FT TRANSMEM 96 116 POTENTIAL.

```
FT TRANSMEM 119 POTENTIAL.
FT TRANSMEM 167 POTENTIAL.
FT TRANSMEM 187 POTENTIAL.
FT TRANSMEM 264 POTENTIAL.
FT TRANSMEM 284 POTENTIAL.
FT TRANSMEM 299 POTENTIAL.
FT TRANSMEM 327 POTENTIAL.
FT TRANSMEM 347 POTENTIAL.
FT TRANSMEM 350 POTENTIAL.
FT TRANSMEM 370 POTENTIAL.
FT TRANSMEM 397 POTENTIAL.
FT TRANSMEM 417 POTENTIAL.
FT TRANSMEM 438 POTENTIAL.
SQ SEQUENCE 452 AA; 50318 MW; 5A596DEA229205C2 CRC64;

Query Match
Best Local Similarity 8.9%; Score 255.5; DB 1; Length 452;
Matches 114; Conservative 92; Mismatches 160; Indels 109; Gaps 24;

QY 84 YRLATIANFALGEGROVDLQLEQESCLDGM-FESQDVYLSITVTE-WNLVCEDDWK 141
DB 21 YRIFGIISFLL-----LTGFLSYSGNVVLAKLVSNWS---NNFLN 59

QY 142 APLTISLFFVGLGSPISQSLDRFRGK-----NVLFVTMGMTQFSLQIFSKNFEMP 196
DB 60 APTSALMF-GYFISLGTGFGIDYFGRRAFRINLLVIGIA-ATGAAFVP-----DMY 111

QY 197 VWLF-----VLVGMGOISNVAAFLVLTGTEILGKSVRIIF-STLGVCIYVAFGY 243
DB 112 WLIFFRFLMTGTMGALIMGVASFTEIFATYRGK---WSARLSFVGNWSPMLSAAGV 167

QY 244 MVLPFLAFYFRDRMMLVLTALVTPGVLCVAMWFP-----ESPRMLISOGREAEVIR 298
DB 168 VWIAFVS-----WRIMFL-----GGIGILLAWFLSGKYFIESPRWLAGQIAGAEQCLR 218

QY 299 KA-----AKANGIVVPSTIFDPSELQDLSKSKQOSHNLIDLLRTWIRMTIMSILMWTI 354
DB 219 EVESQIEREKSLP-----PLTSYOSNKVKVIGKTFWLLFKGEMRLTLVAITVLIAM 273

QY 355 SVGYFGLSLDTPNLHGDIIVN-----CFLSAMVEVPAYV---LAWLLLOYLPRRYSM 403
DB 274 NISLYTITWIP-----IFVNSGIDVDKSLMTAVIMICAPVGIETALIIDHFPR--- 326

QY 404 ATALFGLGSLVLLPMLVPPDLYLATVLMVGVKFGVTAAFSM-VY-----VYTAELPT 456
DB 327 -----LFGSLLLIITAVLYVYSIQTETWAILIYGLVMIFFLYMVCFASAYIPELWPT 381

QY 457 VVRNMGVGSYSTASRLGSLSPFV---YLGAYDRFLPYILMGSLILTA-ILTLF 508
DB 382 HLRGSGFVNAVGRIVAVFTPYGVAALLTHYGSITVFMVGLVMILLCALVLSIF 436
```

```
RESULT 15
PCAK_ACICA STANDARD; PRT; 457 AA.
AC Q43975;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE 4-HYDROXYBENZONATE TRANSPORTER.
GN PCAK.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxId=471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BD413 / ADP1;
RX MEDLINE=94341565; PubMed=8063101;
RA Kowalchuk G.A., Hartnett G.B., Benson A., Houghton J.E., Ngai K.-L.,
RA Ornstom L.N.;
RT "Contrasting patterns of evolutionary divergence within the
RT Acinetobacter calcoaceticus pca operon."
RL Gene 146:23-30(1994).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L05770; AAC37151.1; -.
CC InterPro: IPR003662; sub_transportr.
CC Pfam: PF00083; sugar_tr; 1.
CC PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
CC PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
CC KW transport; Transmembrane; Inner membrane.
CC FT TRANSMEM 35 55 POTENTIAL.
CC FT TRANSMEM 73 93 POTENTIAL.
CC FT TRANSMEM 102 122 POTENTIAL.
CC FT TRANSMEM 129 149 POTENTIAL.
CC FT TRANSMEM 169 189 POTENTIAL.
CC FT TRANSMEM 192 212 POTENTIAL.
CC FT TRANSMEM 275 295 POTENTIAL.
CC FT TRANSMEM 311 331 POTENTIAL.
CC FT TRANSMEM 339 359 POTENTIAL.
CC FT TRANSMEM 365 385 POTENTIAL.
CC FT TRANSMEM 401 421 POTENTIAL.
CC FT TRANSMEM 427 447 POTENTIAL.
CC SQ SEQUENCE 457 AA; 49277 MW; 4F5B5F77361A1567 CRC64;
```

```
Query Match
Best Local Similarity 8.8%; Score 252.5; DB 1; Length 457;
Matches 113; Conservative 78; Mismatches 177; Indels 85; Gaps 20;

QY 110 QESCLDQNEFSQDVYLSITVTEW---NLVC-----EDDW---KAP 143
DB 15 QRSSLDQAALINDAPLRY---QWLIIVCFIVFDGIDTAAMGFIAPALAAQMGVDRSQ 72

QY 144 L--TISLFFVGLGSPISQSLDRFRGKKNVLFVTMGMTQFSLQIFSKNFEMFVVLV 201
DB 73 LGPVNSAALGMIIGALVSGPTDRFRGKIVLSMSMLVFGFTLACAYSTNLSLVIFR 132

QY 202 LVGMGOISNVAAFLVLTGTEILGKSVRIIFSTLGVCIYVAFGYVWLPFLAFYFIRDW----- 256
DB 133 LTGIGLGAAMPNATLTFSEYCPARIR---SLLVTCMF---CGYNLGMAGGFISSWLIPAF 187

QY 257 ---RMLLVALTMPGVLCAVLMWFIPESPRMLISQGR-FEEAEVIIRKAA--KANGIVVPS 310
DB 188 GWSLFLGLGGWAPLILMLLVIFLFPESYRFLIVKGNKTKVROILSRIAPQKVGQV---T 244

QY 311 TIFDPSLQDLSKK-----QQSHNILDLLRTWIRMTIMSILMWTISVGYFGL 361
DB 245 EFHVPEEKVEAGTKKGVFGMLFSKAYVYKGVTLVWTVYFMGLVMYILTLTSLMPTLMRETGA 304

QY 362 SLDPNPLHGDIF-VNCFLSAMVEVPVAVLAWLLLOYLPR----YSNATAFLGSGVLLF 416
DB 305 SLERAARFLGFLQFGGVLSAL-----FIGWAMDRNPNNRIAGFYLAACIF---AVVG 355

QY 417 MQLVPPDLYLATVLMVGVKFGVTAAFSMVYVYTAELPTVTVRNMGVGSYSTASRLGSL 476
DB 356 QSLSNPTL--LALFILCAG-IAVNGAQSSMPVLSARFYPTQCRATGVANMSGIGREGAV- 411

QY 477 SPYFVILGAYDRFLPYILMGSLILTAITLFL 509
DB 412 --FGAWTGA-----VLLGNNSFTMLISMLI 435
```

Search completed: January 25, 2002, 10:15:21  
Job time: 327 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2002, 10:14:54 ; Search time 47.78 Seconds  
(without alignments)  
1705.184 Million cell upd

Title: US-09-521-195-3  
 perfect score: 2883  
 Sequence: 1 MRDYDEVTAFLGSGPFQRL.....HTRMLKDGQERPTILKSTAF 557

Scoring table: BLOSUM62

scoring table:  
 0.000000  
 Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database :

```

SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_protein:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2295.5	79.6	564	11	Q9WNT6	Q9wt66 mus musculus
2	2227	77.2	551	4	Q9H015	Q9h015 homo sapien
3	2218	76.9	551	4	O14546	O14546 homo sapien
4	2157	74.8	553	11	Q9R141	Q9r141 rattus norv
5	2148	74.5	553	11	Q9Z306	Q9z306 mus musculus
6	885	30.7	548	5	Q9VC42	Q9vca2 drosophila
7	881	30.6	548	5	O01384	O01384 drosophila
8	816	28.3	567	5	Q9VCA3	Q9vca3 drosophila
9	777	27.0	568	5	Q9V539	Q9v539 caenorhabdi
10	777	27.0	576	5	O02270	O02270 caenorhabdi
11	753.5	26.1	561	5	Q9V616	Q9v616 drosophila
12	740	25.7	554	4	Q9NQD4	Q9ngd4 homo sapien
13	738.5	25.6	556	4	O75751	O75751 homo sapien
14	738	25.6	554	4	O15395	O15395 homo sapien
15	734	25.5	554	4	O15245	O15245 homo sapien
16	733	25.4	593	11	P70485	P70485 rattus norv
17	728	25.3	593	11	P97558	P97558 rattus norv
18	727	25.2	555	4	O15244	O15244 homo sapien
19	724	25.1	555	11	Q9R0W2	Q9r0w2 rattus norv

## ALIGNMENTS

```

RESULT      1
Q9WTN6      PRELIMINARY;          PRT;       564 AA.
Q9WTN6;
AC          Q9WTN6;
DT          01-NOV-1999 (TrEMBLrel. 12, Created)
DT          01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT          01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
OC          OCTN3.
GN          SLC22A9 OR OCTN3.
OS          Mus musculus (Mouse).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus..
OX          NCBI_TaxID=10090;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          TISSUE=TESTIS;
RA          Nezu J.;
RT          "Mouse OCTN3 - a novel OCTN transporter family protein.";
RL          Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC          -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC          -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR          EMBL; AB018436; BAA78343.1; -.
DR          MGD; MGI:1929481; SLC22a9.
DR          InterPro; IPR003662; sub.transporter.
DR          Pfam; PF00083; sugar.tr.1.
DR          PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW          Transmembrane.
SQ          SEQUENCE      564 AA;   63320 MW;   C37FDCAC6395DAD01 CRC64;

Query Match              79.6%; Score 2295.5; DB 11; Length 564;
Best Local Similarity    80.6%; Pred. No. 7.5e-136;
Matches 439; Conservative 54; Mismatches 48; Indels 3; Gaps 1

Qy      1 MRDYDEVTAFLGEGWGFQRLIFLLSASIIPIINGFTGLSSVFELIATPEHRCRVPDAANLSS 60
        |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      1 MLDYDEVTAFLGEGWGFQRLIFLLSASIIPIINGFTGLSAVELTALPEHRCRIPTVNLS 60

Qy      61 AWRNHVVPLRLRGREVPHSICRRYLATIANFSALGLEPGRDVDLGQLQEBCSLDGWEFS 120
        |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      61 AWRNHSPMETKGGVEPQPKCRRYLATIANFSSELGLEPGRDVDLEQEQENCILDGWED 120

```



```
QY 181 TGFSELOIFSKNFEMFVLVFLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCIFFA 240
DB 181 TGFSELOIFSKNFEMFVLVFLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCIFFA 240
QY 241 FGYMVLPLFAFYFIRDRWMLLVALTMPGVLCVWLFIPESPRWLLISQGRFEAEVIRKA 300
DB 241 VGYMLPLFAFYFIRDRWMLLVALTMPGVLCVWLFIPESPRWLLISQGRFEAEVIRKA 300
QY 301 AKANGIVVPTIFDPSELQDSSKKQSHNLDLRTNIRMTVIMSIMLWMTISVGYFG 360
DB 301 AKMNTAVPAVIFD--SVEELNPLKQKRAFILDFTFRNIAIMTMSLLMLTSGVYFA 358
QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLILLOYLPRYSMATALFLGSGVLLFMOLV 420
DB 359 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLILLOYLPRYSMATALFLGSGVLLFMOLV 418
QY 421 PPDLYLATVLMVGMKFGVTAAFSMVYVYTAELYPTVVRNMGVGVSTASRLGSLSPYF 480
DB 419 PVDYFELSLGLVLMGKFGVTAAFSMVYVYTAELYPTVVRNMGVGVSTASRLGSLSPYF 478
QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVGMKHKRTPSHTR 540
DB 479 VYLGAYNRMLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVGMKHKRTPSHTR 535
QY 541 MKDGOERPTILKSTAF 557
DB 536 DSMTEENPKVL-ITAF 551

RESULT 4
Q9R141 ID Q9R141 PRELIMINARY; PRT; 553 AA.
AC Q9R141;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ORGANIC CATION TRANSPORTER OCTN1.
GN OCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu X., Wang H., Leibach F.H., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of OCTN1,
an organic cation transporter."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF169831; AAD46922.1; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 553 AA; 62362 MW; E26C8155768A14AD CRC64;

Query Match 74.8%; Score 2157; DB 11; Length 553;
Best Local Similarity 72.7%; Pred. No. 3.3e-127;
Matches 405; Conservative 65; Mismatches 83; Indels 4; Gaps 2;

QY 1 MRDYDEVTAFLGEGWGFQRLIFLFLSASIIPIGTGLSSVFLIATPEHRCRVPDANLSS 60
DB 1 MRDYDEVIAFLGEGWGFQRLIFLFLSASIIPIGTGLSSVFLIATPEHRCRVPDANLSS 60
QY 61 AWRNHTVPLRLRDGVRPHSCRRYRLATIANFSALGLEPGRVDLQLEQESCLDGWEYS 120
DB 61 AWRNHTVPLRLRDGVRPHSCRRYRLATIANFSALGLEPGRVDLQLEQESCLDGWEYS 120
QY 121 QDVLSTIVTEWNLVCEDDWKPTLTTSLFPVGLGSEFVSQGLSDRFRGKKVLFATMAVQ 180
DB 121 QDVLSTIVTEWNLVCEDDWKPTLTTSLFPVGLGSEFVSQGLSDRFRGKKVLFATMAVQ 180
```

```
DB 121 QDVLSTIVTEWNLVCEDDWKPTLTTSLFPVGLGSEFVSQGLSDRFRGKKVLFATMAVQ 180
QY 181 TGFSELOIFSKNFEMFVLVFLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCIFFA 240
DB 181 TGFSELOIFSKNFEMFVLVFLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCIFFA 240
QY 241 FGYMVLPLFAFYFIRDRWMLLVALTMPGVLCVWLFIPESPRWLLISQGRFEAEVIRKA 300
DB 241 VGYMLPLFAFYFIRDRWMLLVALTMPGVLCVWLFIPESPRWLLISQGRFEAEVIRKA 300
QY 301 AKANGIVVPTIFDPSELQDSSKKQSHNLDLRTNIRMTVIMSIMLWMTISVGYFG 360
DB 301 AKMNGIMAPAVIFDPELQELNSLKQOKVFLDLFKTRNIATITVMSVLMWMLTSGVYFA 360
QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLILLOYLPRYSMATALFLGSGVLLFMOLV 420
DB 361 LSLNPNLHGDVYLCFLSLGIEVPAYFTAWLLRTLPRIIAGVLFWGGVLLLVQVV 420
QY 421 PPDLYLATVLMVGMKFGVTAAFSMVYVYTAELYPTVVRNMGVGVSTASRLGSLSPYF 480
DB 421 PEDYNFVSLGLVLMGKFGVTAAFSMVYVYTAELYPTVVRNMGVGVSTASRLGSLSPYF 480
QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVGMKHKRTPSHTR 540
DB 481 VYLGAYNRMLPYILMGSLTILTAITLFLPESFGTLPDTIDOMLRVGMKHKRTPSHTR 537
QY 541 MKDGOERPTILKSTAF 557
DB 538 VSMDEENPKVL-ITAF 553

RESULT 5
Q9Z306 ID Q9Z306 PRELIMINARY; PRT; 553 AA.
AC Q9Z306;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ORGANIC CATION TRANSPORTER.
GN SLC22A4 OR OCTN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL 6J; TISSUE=KIDNEY;
RA Nezu J.;
RT "Mouse OCTN1 : Polyspecific organic cation transporter."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AB016257; BAA36626.1; -.
DR MGD; MGI:1353479; SLC22A4.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 553 AA; 62290 MW; C4D66BC061398653 CRC64;

Query Match 74.5%; Score 2148; DB 11; Length 553;
Best Local Similarity 72.2%; Pred. No. 1.2e-126;
Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;

QY 1 MRDYDEVTAFLGEGWGFQRLIFLFLSASIIPIGTGLSSVFLIATPEHRCRVPDANLSS 60
DB 1 MRDYDEVIAFLGEGWGFQRLIFLFLSASIIPIGTGLSSVFLIATPEHRCRVPDANLSS 60
QY 61 AWRNHTVPLRLRDGVRPHSCRRYRLATIANFSALGLEPGRVDLQLEQESCLDGWEYS 120
DB 61 SWRNHSIPLTKDGRVQPSQRRYRLATIANFSAMGLEPQDQVDLEQLEQESCLDGWEYD 120
```

QY 121 ODVYLTIVTEWNLVCEDDWKAPLTLISLFFVGVLLGSFTSGQLSDRFRGKKNVLFVTMGQ 180  
 DB 121 KIDFLSTIVTEWNLVCEDDWKAPLTLISLFFVGVLLGSFTSGQLSDRFRGKKNVLFVTMGQ 180  
 QY 181 TGFSLQIFSKNFEMFVFLVGVGQISNYAAVFLGTEILGKSVRIIFSLGVCIEYA 240  
 DB 181 TGFSLQIFSKNFEMFVFLVGVGQISNYAAVFLGTEILGKSVRIIFSLGVCIEYA 240  
 QY 241 FGYMVLPLFAYFIRDRWMLLVALTMPGVLCVALWFIPESPRWLISQGRFEAEVIRKA 300  
 DB 241 FGYMVLPLFAYFIRDRWMLLVALTMPGVLCVALWFIPESPRWLISQGRFEAEVIRKA 300  
 QY 301 AKANGIWPSTFDPSELQDLKSKQOQSHNIDLTNRWNTMSIMLWMTISGVYFG 360  
 DB 301 AKANGIWPSTFDPSELQDLKSKQOQSHNIDLTNRWNTMSIMLWMTISGVYFG 360  
 QY 361 LSLDTPNLHGDIFVNCFLSAMVEYPAVLAWLLLOVLPYRYMATATLFGGVLVLMQV 420  
 DB 361 LSLDTPNLHGDIFVNCFLSAMVEYPAVLAWLLLOVLPYRYMATATLFGGVLVLMQV 420  
 QY 421 PPDLYLATVLMVWGKFGVTAAFSMVYVYTAELYPTVVRNMGVGVSTASRLGSLSPYF 480  
 DB 421 PPDLYLATVLMVWGKFGVTAAFSMVYVYTAELYPTVVRNMGVGVSTASRLGSLSPYF 480  
 QY 481 VILGAYDRFLPYLWGLSILTLTALTLFLPESFGTFLPDTIDQMLRVKGMKHKRTPSHR 540  
 DB 481 VILGAYDRFLPYLWGLSILTLTALTLFLPESFGTFLPDTIDQMLRVKGMKHKRTPSHR 540  
 QY 541 MLKQGOERPTILKSAF 557  
 DB 541 MLKQGOERPTILKSAF 557  
 QY 538 VSVDRSPKVL-IPAF 553  
 DB 538 VSVDRSPKVL-IPAF 553

RESULT 6  
 Q9VCA2 ID Q9VCA2 PRELIMINARY; PRT: 548 AA.  
 AC Q9VCA2;  
 DT 01-MAY-2000 (TRENBLTel. 13, Created)  
 DT 01-MAY-2000 (TRENBLTel. 13, Last sequence update)  
 DT 01-JUN-2001 (TRENBLTel. 17, Last annotation update)  
 DE ORCT PROTEIN.  
 GN ORCT OR CG6331.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RESULT  
 001384

7

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 DR EMBL; AE003747; AAF56271.1; -;  
 DR FlyBase; FBgn0019952; Orct.  
 DR InterPro; IPR003662; sub.transporter.  
 DR Pfam; PF00083; sugar\_tr; 1.  
 KW Transmembrane.  
 SQ SEQUENCE 548 AA; 61002 MW; 08D7F97599B477AF CRC64;

Query Match 30.7%; Score 885; DB 5; Length 548;  
 Best Local Similarity 35.1%; Pred. No. 9.9e-48;

Matches 198; Conservative 106; Mismatches 206; Indels 54; Gaps 9;

QY 4 YDEVATFLGEGWGPQRLIFLFLLSASIIIPNGFTGLSSVFLIATPEHRCRVP----DAANLS 59  
 DB 3 YDDVITHLGEFGPYKRIYYLLCLPAIVCAFHKLGLAGVFLAKDPFCALPYENGSIYELS 62  
 QY 60 SAWENHTVPLRLDGRVPHSCRRYRLATIANFSLGLEPGRDVL-----CQL----- 108  
 DB 63 PHLMNLSYPNER-----CSYI-----DVDYTEYLNGSIIPRSSN 97  
 QY 109 EQESCLDGEWFEQDQVYLSTIVTEWNLVCEDDWKAPLTLISLFFVGVLLGSFTSGQLSDRFG 168  
 DB 98 ETKTC-SSYVVDYRSKYLNSAVTEWNLVCSRLSLSATSDSLFMLGVLLGSLIFGQMSDKLG 156  
 QY 169 RKNVLFVTMGWQGFSPQLQIFSNFEMFVFLVGVGQISNYAAVFLGTEILGKSVRI 228  
 DB 157 RKPTFFASLVQLIFGVLAIAVAPEYFSTIRMIIVGATTSGVFLVAVIALENVGSYSYR- 215  
 QY 229 IESTLGVCIPIYAFGYMVLPLFAFYFIRDRWMLLVALTMPGVLCVALWFIPESPRWLISOG 288  
 DB 216 LFGAVAMQMFSSVGFMLTAGFYFIRDRWMLLVALTMPGVLCVALWFIPESPRWLISOG 275  
 QY 289 RFEEAEVIRKAANGIVVPSTIFD-----PSELQDLSSKKQOQSHNIDLTNRWIR 341  
 DB 276 RKDEAFVILIEKAENKVEPNIEYQLVDEVAEKKQDEMAASQPAATVFDLLRYPNLR 335  
 QY 342 MYTMSIMLWMTISGVYGLSLDTPNLHGDIFVNCPLSAMVEYPAVLAWLLQVLPYRY 401  
 DB 336 RKTLLIFFDFVFNVSQVYGLSWNTNLGGNQLVNFEMISGAVEIPGYTLLFTLLNRGRRS 395  
 QY 402 SMATALFLGSSVLELQVLPDLYLATVLMVWGKFGVTAAFSMVYVYTAELYPTVVRNM 461  
 DB 396 ILCTGMVAGISLULATIFVPSDMNMLIVACAMIGLUATISSYGTIIFSAEQFPTVVRNV 455  
 QY 462 GVGVSSTASRLGSLSPYFVFLGAYDRFLPYLIMGSLTILTLTALTLFLPESFGTFLPDTI 521  
 DB 456 GLGASSWAVRGVIGILAPYLKLLGTEWRPLIICGALSITAGLLSILLPETLNKWPETI 515  
 QY 522 DQMLRVKGMKHKRTPSHRMLKDG 545  
 DB 516 ED-----GENFGKKPAPQETAEG 534

ID	001384	PRELIMINARY;	PRT;	548 AA.
AC	001384;			
DT	01-JUL-1997	(TrEMBLrel. 04, Created)		
DT	01-JUN-1997	(TrEMBLrel. 04, Last sequence update)		
DT	01-JUL-2001	(TrEMBLrel. 17, Last annotation update)		
DE	PUTATIVE ORGANIC CATION TRANSPORTER.			
GN	ORCT OR CG6331.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxId=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98072431; PubMed=9409773;			
RA	Taylor C.A.M., Stanley K., Shirras A.D.;			
RT	"The Orct gene of Drosophila melanogaster codes for a putative organic			
RT	cation transporter with six or 12 transmembrane domains.";			
RL	Gene 201:69-74(1997).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).			
CC	-1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.			
DR	EMBL; Y12400; CAAY3031.1; -			
DR	FlyBase; FBgn0019952; Orct.			
DR	InterPro; IPR003662; Sub. transporter.			
DR	Pfam; PF00083; sugar_tr; 1.			
DR	Transmembrane.			
KW	SEQUENCE 548 AA; 61105 MW; 2665B71C969C5E4A CRC64;			
Query Match 30.6%; Score 881; DB 5; Length 548;				
Best Local Similarity 34.9%; Pred No. 1.8e-47;				
Matches 197; Conservative 107; Mismatches 206; Indels 54; Gaps				
Qy	4	YDEVTAFLGEMGFPQRLIFFLLSASITPNFGTGLSSVFLIATPEHRCRVP-----DAANLS	59	
Db	3	YDDVITHLGFEGPYQKRIYYLLCLPAIVCAFHKLAGVFLAKPDRFCALPYENGSIYELS	62	
Qy	60	SAWENHNVPLRLDGRVPHSCRRYRLATIANFSALGLEPGROVDL-----GQL-----	108	
Db	63	PHLNLSPENER-----CSY-----DVDITEYLNSTIPRSSN	97	
Qy	109	BQESCLDGEFSQDVLSTIVTWNLCEDDWMKAPLTIISFFVGVLGSLFSGQLSDRFG	168	
Db	98	ETKTC-SSYVDRSKYLSNAVTEWNLVCSRLLSATSDSLFLGVLLGSLIFGQMSDKLG	156	
Qy	169	RKNVLFVTMGQGTGFLQFSKNFEMFVFLVFLVGMGQISNVAFAVFLGTHILCKSVRI	228	
Db	157	RKPTFFASVLQLIFGLAAVAPEYFSYTSIRNIVGATTSGVFLVAVLEAMGVGSSYR-	215	
Qy	229	IFSTLGVCFYAFGYVYVLPFAFYFIDRWMLLVALMPGVLCVLAWVWFIPESRWLISQ	288	
Db	216	LPAGVAMQMFSGVGMLTAGFAFIHWRWLQIATILPGLFLCYWIIPESARWLLMKG	275	
Qy	289	RFEAEVIRKAAGKANGVVPSTFD-----PSELQDLSSKKOOSHNLDLLRLTNWIR	341	
Db	276	RKDEAFVIEKAANKENKVEPNETIEQLVDEVAEKKKQDEMAASQAPATVFOLLRYPNLR	335	
Qy	342	MVTNLSMLNMTISGVFGSLDTPNLHGDFIVNCFLSAMVEVPAIVLAWLLQVLPTRY	401	
Db	336	RKTLILFIDFVNVSGVYVLSWNTNLGGNQLNFMISGEVPIGTYLLEFFTLNRWGRS	395	
Qy	402	SNATALFLGGSVLLFMQLVPPDLVYATLVLVWVKFGVTAAFSMVYVYFAELYPTWRNM	461	
Db	396	ILCGTMVAGISLLATITFVPSDMNLIVACAMTGLKLAITSYSGTIVIFSAGQFPVWRNV	455	
Qy	462	GVGVSSTASRLGSILSPYFVYLGA YDRFLPYILMGSLLTILTLFLPESFGTLPDPTI	521	
Db	456	GLGASSMVARVGGILAPYLKLLGEIWRPLPLIICGALS TAGLLSRLLPETLNKMPETI	515	
Qy	522	DOMLEVGMKIRKTPSHTRMLKDG	545	
Db	516	ED-----GENFGKKPAQETAEEG	534	

```

RESULT      8
Q9VCA3
ID          PRELIMINARY;      PRT;      567 AA.
AC          Q9VCA3;
DT          01-MAY-2000 (TEMBLrel. 13, Created)
DT          01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT          01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE          CGI3610 PROTEIN.
GN          CGI3610.
OS          Drosophila melanogaster (Fruit fly).
OC          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC          Pterygota; Neoptera; Endopterygota; Diptera; Brachyoptera; Muscomorpha;
OC          Ephydroidea; Drosophilidae; Drosophila.
RN          NCBI_TaxID=7227;
RX          [1]
SEQUENCE FROM N.A.
RC          STRAIN=BERKELEY;
RX          MEDLINE=20196006; PubMed=10731132;
RA          Adams M.D., Celikler S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA          Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA          George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA          Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA          Brandon R.G., Rogers Y.-H.C., Blazej R.G., Champe M.R., Pfeiffer B.D.,
RA          Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA          Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA          Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA          Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA          Borokva D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA          Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA          Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA          de Pablos B., Delcher A., Deng Z., Duan A.D., Dew I., Dietz S.M.,
RA          Dodson K., Doup L.E., Downes M., Dugan-Rochet S., Dunkov B.C., Dunn P.,
RA          Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA          Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA          Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA          Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA          Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA          Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA          Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA          Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA          Liu X., Mattet B., McIntosh A.C., McLeod M.P., McPherson D.,
RA          Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA          Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA          Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA          Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
RA          Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA          Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA          Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA          Svarks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA          Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA          Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA          Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA          Zheng X.H., Zhong F., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA          Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT          "The genome sequence of Drosophila melanogaster.";
RL          Science 287:2185-2195(2000).
CC          -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC          -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR          EMBL; AE003747; AAF56270.1; -
DR          FlyBase; FBgn0039176; CGI3610.
DR          InterPro; IPR003662; sub_transporter.
DR          Pfam; PF00083; sugar_tr; 1.
KW          Transmembrane.
SQ          SEQUENCE      567 AA;      63166 MW;      0D9B979469395E60 CRC64;

```

Query Match	28.3%;	Score 816;	DB 5;	Length 567;
Best Local Similarity	32.2%;	Pred. No. 2.1e-43;		
Matches 183:	Conservative 121;	Mismatches 212;	Indels 52;	Gaps 9

QY 4 YDEVTAFLGEWGPQRLIFLLSASIIPNGFTGLSSVFLIATPEHRCRYPDAANLSSAWR 63



```

Db 3 YDEAIHLGDFGRYQKIIFELICLTSIPVAFHKLAGVFLAKDPDFRCALPFENGSSYDLP 62
QY 64 NHTVPLRLRDGREGVPHSCRRYRLATIANFSALEGLPEGRDVL-----GQL-----EDES 112
Db 63 THLWNLSPENER-----CSY-----DVTTEYLNGSPRPSNETKT 101
QY 113 CLDWGFEQDVLSTIVTWNLCEDDKAPLTISLFFVGVLLGSFISQSLSDREGRKNV 172
Db 102 C-SSYVYDRSKYLSAVTEWNLVCGRDPWAATSDSLFMLGVLLGSIVFGQLSDKYGRKPI 160
QY 173 LFTWGMQTFGLFQIFSKNFEMFVFLVLCMGQISNVAAVFLGTEILGKSVRIIFST 232
Db 161 LFASLVIOQLFGLVAGVAPEYFTYTFARLMVGATTSGVFLVAVVAMEMVGPDKR-LYAG 219
QY 233 LGVCIFYAFGYVPLPFAVFIQWRMLLVALTMPGVLCVLMWETPESPRWLISOGREE 292
Db 220 IFWMEFFSGVMLTAVFAVFDHWRQLALPLGLIFWYIILPESARWLLKGRDC 279
QY 293 AEVIIRKAAKANGIVP-----STFDPSF-----LQDLSKKQKOSHNLIDLLRTWN 339
Db 280 ATANMOKAARFNKVEISDEALSELLDEGENSEKAKQLEQDELDEGPPPSWMDLFCYPN 339
QY 340 IRWVTIMSLWMTISVGVGLSLDTPNLHGDIYVNCFLSAMVEPAYVLAWLILQYLPR 399
Db 340 LRKTLILFLDLVTSVGYVGLSWNTSLGGNVLLNFVISGAVEIPAYIFILLTLNRWR 399
QY 400 RYSMATALFGLGSVLELMQVPPDLYLATVLMVVGKFGVTAAFSMVYVYTAELYPTVVR 459
Db 400 RSLGCLVAGLSLATVLIIPQRMHTLVACAMLGKLAITASYGTVIFSAEQFPTVVR 459
QY 460 NMGVGSVSTASRLGSLSPYFVLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPD 519
Db 460 NYALGAASVARISSQMAPFLNATLWKPPLLCGLSLTIVAGLLSLLLPETHNKPML 519
QY 520 TTDQMLRVKGMKHKRTPSHTRMLKDGQE 547
Db 520 TI-----ADGERFGKTRADVLETGQE 542

```

```

Query Match 27.08; Score 777; DB 5; Length 568;
Best Local Similarity 31.4%; Pred. No. 5.8e-41;
Matches 165; Conservative 118; Mismatches 225; Indels 18; Gaps 6;
QY 3 DYDEVTAFLGEGWGFQRLIFFLLS-ASIIIPNGFTGLSVFLIATPEHRCRVPDAANLSSA 61

```

```

Db 8 DFDVFLQVGVNGYQIVYVFELICLPTSLPSAFSAFNIPFVGNPPHTCHIPGKEYLRP 67
QY 62 WRNITVPLRLRDGREGVPHSCRRYRLATIANFSALEGLPEGRDVLQGLEQSLDGEFQ 121
Db 68 LTNDTQIL-----SCKQYNETQINVFRAFTSAP-VDTYSDRISLVPCCQNGWDYDN 116
QY 122 DVYLSTIVTWNLCEDDKAPLTISLFFVGVLLGSFISQSLSDREGRKNVLFVWGMQTF 181
Db 117 STYLDLSLVTEFNLVCDQQAWEISTTSFYVGSFIGNCLFGYVADKFGRRRSFFVILT 176
QY 182 GFSEFLQIFSKNFEMFVFLVLCMGQISNVAAVFLGTEILGKSVRIIFSTGLGCIYAF 241
Db 177 VCGFASFAKDIESFIILREFTGLAPALQIPFIICMEFGNSGR-IFSGLMSTLFFGA 235
QY 242 GYVVLPLFAVFIQWRMLLVALTMPGVLCVLMWETPESPRWLISOGREEAEVIRKAA 301
Db 236 AMALLGVVAMFIRWRQLTFFCNAPFAFYIYFFLPESPRWSVSGKWADAKKOLKKA 295
QY 302 KANG--IVVPSTIFDPSLQDLSKKQ--OSHNLIDLLRTWIRWVTIMSLWMTISV 356
Db 296 KMGKSNVDDELVDLSMKNHONAAEKETARSHNVTDLPKTPNLRRKTLIVTIWVWNAI 355
QY 357 GYFGLSLDTPNLHGDIYVNCFLSAMVEPAYVLAWLILQYLPRYSMATALFGLGSVLLF 416
Db 356 IYNGLTENSVNLPVDDYWSFIINGAVELPGYFVVMVPLLCAGRRWTLAATMIVCGIGCVS 415
QY 417 MOLVPPDLYLATVLMVVGKFGVTAAFSMVYVYTAELYPTVVRNMGVGSVSTASRLGSL 476
Db 416 AMFMDGYPMLVASASFIKFGVSGFAVIYIFAGELYPTVVRNMGVGSVSTASRLGSL 475
QY 477 SPYFVYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDITD 522
Db 476 APHVNLGKIVKILPULLIMGLMALSAGILFFLPETLGAPLPMTIE 521

```

RESULT 10  
O02270 PRELIMINARY; PRT; 576 AA.

AC O02270;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE F52F12.1 PROTEIN.  
GN F52F12.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
Raxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,  
Granger A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,  
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P., III of C.  
RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.;  
RL Nature 368:32-38(1994).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL; Z83228; CAB05732.1; -;  
DR InterPro; IPR003662; sub\_transporter.



```
DR Pfam: PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 576 AA; 64493 MW; F494EE94A7EBC0B1 CRC64;

Query Match 27.0%; Score 777; DB 5; Length 576;
Best Local Similarity 31.4%; Pred. No. 5.9e-41;
Matches 165; Conservative 118; Mismatches 225; Indels 18; Gaps 6;

QY 3 DYDEVTAFLGEGWGFQRLIFLLS-ASIPNGTGLSSVFLIATPEHRCRVDPDAANLSA 61
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 16 DFEVLEQVQNGYGYIVFFIICLTSLPSAFSAPFNIPFVGNPHTCHIEPKREYLRP 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 WRNHTVRLRLDRGVEPHSCRRYRLATIANFSALEGPGRDVLGOLESCLDGEWFSQ 121
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 76 LTNDTQL-----SKQNETQINVFRAFTSAP-VDTYSDRISLVPQNGMDYDN 124

QY 122 DVLSTIVTEWNLVCEDDMKAPLTISLFFYGVLLGSFISQSLSDRGKRNVLFTVMGMQT 181
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 125 STYLDLSLVTENLVDCQQAWEISTTSFYVGSFIGNCLFGYVADKGRRRSPFVILTILI 184

QY 182 GFSPLQIFSNNFMFVFLVFLVGMQISNVAAVFLGTEILGKSVRLIISTLGCVFYAF 241
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 185 VCGTASSFAKDIESFIILRFTTGLAPALFQIPFIICMEFGNSGR-IFSGLMTSLFFGA 243

QY 242 GYVLPFLFAYFIRWRMLLVALTMPGVLCVALWFIPESPRWLISOGREPEAEVIIRKAA 301
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 244 AMALLGVAMFIRKWRQLTFFCNAPPAFYIYIFFLPESPRWSVGVKNADAKKQLKKIA 303

QY 302 KANG---IVVPSTIFDSELDLSSKKQ--QSHNLDLRLTNRIMVTIMSLMWTISV 356
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 304 KMGKSNVDVDELVDSMKNHONAAEETKRSHNVTDLFKTPNLRKRTLIVTIWVNAI 363

QY 357 GYFGLSDTPNLHGDIFVNCFLSNVEVPAYVLAWLLQLLYLPRYSNATLFLGGSVILF 416
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 364 IYNGLTNLVSNLPLVDDYWSFTINGAVELPGYFVVMPLLCACGRRTLAATMIYCGIGCVS 423

QY 417 MOLVPPDLIYLATVLVWVGKFGYTAAFSMVYVTAELVPTVVRNMVGVGSTASRLGSIL 476
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 424 AMFMPDGYPLVASASFIGKGVGSGFAYIYFAGELYPTVVRNIGMSWVAGSGLLL 483

QY 477 SPYFVYLGAYDRFLPYILMGSLLTILTLTLFLPESFGTLPPTID 522
   I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 484 APHIVNLGKIVKILPLLIMGLMALSAGILTFELPGLGAPLPTIE 529

RESULT 11
Q9V6L6 PRELIMINARY; PRT; 561 AA.
AC Q9V6L6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG4630 PROTEIN.
GN CG4630.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Surtan G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
```

Db 420 GYSLAWLFLRRFRRRVALSGSLLLCSITCVASGFTVLGANLWLVVTLFLVGLKIGITSSRAV 479

QY 446 VYVYTAELPTVVRNMGVSGVSTASRLGSLSPYFVYLGAYDRFLPYILMGSLTILTAIL 505

Db 480 IYFTTAEEMPTVIRSGVGVMSTFARFGAMLAPFVPLLASYYDPLFLLLFGTSLVAGIL 539

QY 506 TFLPESFGTLPPLDT 520

Db 540 SLLLPFTFNKLPDT 554

RESULT 12

Q9NQD4 PRELIMINARY; PRT; 554 AA.

AC Q9NQD4

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE PARTIAL OCT1 ORGANIC CATION TRANSPORTER, EXON 1 AND JOINED CDS.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Hayer M., Bonisch H., Bruss M.;

RT "Molecular cloning, functional characterization and genomic

RT organization of four alternatively spliced isoforms of the human

RT organic cation transporter 1 (hOCT1/SLC22A1).";

RL Ann. Hum. Genet. 63:473-482(1999).

RN [2]

RP SEQUENCE FROM N.A.

RA Hayer M., Bonisch H., Bruss M.;

RT "Molecular cloning, functional characterization and genomic

RT organization of four alternatively spliced isoforms of the human

RT organic cation transporter 1 (hOCT1/SLC22A1) (corrigendum).";

RL Ann. Hum. Genet. 64:267-267(2000).

CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL; AJ243995; CAB95971.1; JOINED.

DR EMBL; AJ243996; CAB95971.1; JOINED.

DR EMBL; AJ276051; CAB95971.1; JOINED.

DR EMBL; AJ276052; CAB95971.1; JOINED.

DR EMBL; AJ276053; CAB95971.1; JOINED.

DR EMBL; AJ245460; CAB95971.1; JOINED.

DR EMBL; AJ243998; CAB95971.1; JOINED.

DR EMBL; AJ243999; CAB95971.1; JOINED.

DR EMBL; AJ244000; CAB95971.1; JOINED.

DR InterPro; IPR003662; sub-transporter.

DR Pfam; PF00083; sugar\_tr; 1.

DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_2.

KW Transmembrane.

SQ SEQUENCE 554 AA; 61169 MW; 9CA1A6C0FC2B2A96 CRC64;

Query Match 25.7%; Score 740; DB 4; Length 554;

Best Local Similarity 35.9%; Pred. No. 1.2e-38;

Matches 197; Conservative 86; Mismatches 215; Indels 50; Gaps 12;

QY 1 MRDYDEVTAFLGEMGPFOR---LIFLLSASTIINGFTGLSSVELIATPEHRCRVDAAN 57

Db 1 MPTVDDILEQVGESGWFKQAFLLICLLSAFAP---ICVGIVELGFTPDHHCOSPGVAE 57

QY 58 LSS---AWR-----NHTVRLRLRDGVEPHSCRRYR-----LATIA-NFSALG 96

Db 58 LSQRGWSPAEELNTVPGLGPAEAFIGQCRRYEVDWQNSALSCVDPPLASLATNRSHLP 117

QY 97 LEPGRDVLGLQEQSLDGNEFSDVYLSITVTEWNLVCEDDWKAPLTISLFFVGVLLG 156

Db 118 Lgp-----CQDGWY---DTPGSSIVTEFNFLVCADSKWLDLFSCLNAGELFG 162

QY 157 SPISQLSDRFGRKVLFTVMGTMQTGFSFLQIFSKNEMFVFLVGMGQISNYAAVF 216

Db 163 SLGVGYFADRFGRKLCLLCTGLTVLVNAVSGVLMAFSPNTMSMLLFRLLQGLVSKGNWAGYT 222

QY 217 LGTEILKSKVRIIPSTLGVCIYAF--GYMVLPLFAYFIRDWRMLLVALTMPGVLCVALW 274

Db 223 LITEFVGSGR--RTVAIMYQMAFTVGLVALTGLAYALPHRWLQLAVSLPTFLFLYY 279

QY 275 WFIPESRWLISQGRFEAEVIRKAAKANGIVVPSIFDPSELODLSKQKQSHNLDL 334

Db 280 WCVPESPRWLLSQKRNTFAIKIMDHIAQKNGKLPADLKLMSLEEDVTER--LSPSFADL 337

QY 335 LRTWNIRMTIMSLMWTISVGVFGLSLDTPNLHGDFVNCFLSAMVPEVAYVLAWL 394

Db 338 FRTLRLKRTFILMFTSVLQGLILHMGATSGNLYLDLYLSALVEIPGAFIALTI 397

QY 395 QYLPRRYMATALFLGGSVLLFMQVPPDYLYLATVLVMVGKFGVTAAFSVVYTAEL 454

Db 398 DRVGRIYPMAMSNLLAGAACLVMIIFISPDHLNLNIIIMCVGRMGITTAIQMICLVNAEL 457

QY 455 PTVVRNMGVGVSTASRLGSLSPYFYV--LGAYDRFLPYILMGSLTILTAILFLPESF 513

Db 458 PTFVRNLGVWVCSSLCDIGGIITPFIIVFRLEVRVQWLPFLFVGLLAAAGVTLILPETK 517

QY 514 GTPLPDTI 521

Db 518 GVALPETM 525

RESULT 13

Q75751 PRELIMINARY; PRT; 556 AA.

AC Q75751; Q9UP02;

DT 01-NOV-1998 (TReMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)

DE ORGANIC CATION TRANSPORTER 3 (EXTRANEURONAL MONOAMINE TRANSPORTER)

DE (EMT) (SOLUTE CARRIER FAMILY 22, MEMBER 3).

GN SLC22A3 OR EMT.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.

RC TISSUE=KIDNEY;

RX MEDLINE=99212254; PubMed=10196521;

RA Gruendemann D., Schechinger B., Rappold G.A., Schoenig E.;

RT "Molecular identification of the corticosterone-sensitive

RT extraneuronal catecholamine transporter.";

RL Nat. Neurosci. 1:349-351(1998).

RN [2]

RP SEQUENCE OF 513-556 FROM N.A., AND TISSUE SPECIFICITY.

RX MEDLINE=99134299; PubMed=9933568;

RA Verhaagh S., Schweifer N., Barlow D.P., Zwart R.;

RT "Cloning of the mouse and human solute carrier 22a3 (SLC22a3/SLC22A3)

RT identifies a conserved cluster of three organic cation transporters on

RT mouse chromosome 17 and human 6q26-q27.";

RL Genomics 55:209-218(1999).

RN [3]

RP FUNCTION, AND TISSUE SPECIFICITY.

RC TISSUE=KIDNEY;

RX MEDLINE=20425388; PubMed=10966924;

RA Wu X., Huang W., Ganapathy M.E., Wang H., Kekuda R., Conway S.J.,

RA Leibach F.H., Ganapathy V.;

RT "Structure, function, and regional distribution of the organic cation

RT transporter OCT3 in the kidney.";

RL Am. J. Physiol. 279:F449-F458(2000).

CC -!- FUNCTION: MEDIATES POTENTIAL-DEPENDENT TRANSPORT OF A VARIETY OF

CC ORGANIC CATIONS.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -!- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, SKELETAL MUSCLE,

CC PROSTATE, AORTA, LIVER, FETAL LUNG, SALIVARY GLAND, ADRENAL GLAND,

CC KIDNEY AND BRAIN CORTEX. NO EXPRESSION DETECTED IN SPLEEN.

CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC

```
CC CATION SUBFAMILY.
DR EMBL; AJ001417; CA004751.1; -.
DR EMBL; AF078749; AAD20977.1; -.
DR MIM; 604842; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW Transport; Ion transport; Transmembrane; Glycoprotein.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 468 488 POTENTIAL.
FT TRANSMEM 498 518 POTENTIAL.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 556 AA; 61279 MW; C3CA2D77DD21C658 CRC64;

Query Match 25.6%; Score 738.5; DB 4; Length 556;
Best Local Similarity 33.0%; Pred. No. 1.4e-38;
Matches 189; Conservative 117; Mismatches 214; Indels 53; Gaps 18;

QY 1 MRDYDEVTAFLEGGPQRLIFELLSASIPNGFTGLSSVFLIATPEHR-CRYPDAANLS 59
DB 1 MPSDEALQVRGFRGFRVLLCLLTGTFAFLVGVVFLGTQPDHYWCGRPSAAALA 60
QY 60 S-AWR-----NHTVPLRLRDGREVPH---SCRRYRLATIANFSA-----LGLE 98
DB 61 ERCGWSPEEWNTAP--ASRGPEPERRGRCQRY--LLEANDSASATLSACADPLAF 117
QY 99 PGRDVLQGLEQESCLDGEFSQDVYLSITVTENLVNLCEDDKWAKPLTISLFFVGVLLG 158
DB 118 PNRSAFL-----VPCRGGWRYAQ--AHSTIVSEFDLCVNVANMLDLTQAILNLGFLTGF 170
QY 159 ISGQLDRFRGKRVLFVTVMQMTGFSFLQI-FSKNFEMFVFLVGMGQISNYVAAFLV 217
DB 171 TLGYAADRYGR-IVYLLSLGIVGTGVVVAFAFPFVFIFRFLQGVFGKGTWTCYVI 229
QY 218 GTEILGKSVRIIFSTLGVCI--FVAFGYMVLPLFAFYFIRDRMMLLVALTMPGVLVALW 275
DB 230 VTELVGSKQRI---VGIVIQMFTEIIGIIPGIAIFIPNWQIQIAITLPSFLFLYIW 286
QY 276 FIPESPRWLISQGRFEAEVIRKAANGIVVPSTIFDPSSELDLSSKKQOQSHNILDLL 335
DB 287 VPESPRWLITRKKGKALQILRIAKCNKYLSSNY---SEI-TVTDEVSNPSFLDLV 342
QY 336 RTWNRIMVTMSIMLWMTISVGFGLSLDPNHLGDIIVNCFISAMVEVPAYVLAWLILQ 395
DB 343 RTPQRRKCTILMFAPFETSAVYQGLVMRLGIIGNLYIDFISGVVLPGLALLITIE 402
QY 396 YLPYRYSNATLALFGSVLLFMQLVPPDLYLATLVLMVCKFGVTAASVYVYTAELVP 455
DB 403 RLGRRLPFAASNVAGVACLVTAFLPGIAWLRTVTATIGRLGITWAFELIVLVNSLXP 462
QY 456 TVVRNMGVSVSTASRLSGSILSPYFY--LGAYDRFLPYILMGSLTITAILTLFLPESFG 514
DB 463 TTLNFGVSLCSGLCDFGGIIAPFLFLRLAAVWLELPLIIFGILASICGGLVMLLPETKG 522
QY 515 TPLPDTIDQMLRV-----KGMKHKRTP-----SH 538
DB 523 IALPETVDDVEKLGSPHSCRCGRNKKTPVSRSH 555

RESULT 14
ID O15395 PRELIMINARY; PRT; 554 AA.
AC O15395;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
```

```
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ORGANIC CATION TRANSPORTER 1.
GN HOCT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97330822; PubMed=9187257;
RA Zhang L., Dresser M.J., Gray A.T., Yost S.C., Terashita S.,
RA Giacomini K.M.;
RT "Cloning and functional expression of a human liver organic cation
RT transporter";
RL Mol. Pharmacol. 51:913-921(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; U77086; AAB67703.1; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW Transmembrane.
SQ SEQUENCE 554 AA; 61153 MW; 55206B897DE32202 CRC64;

Query Match 25.6%; Score 738; DB 4; Length 554;
Best Local Similarity 35.9%; Pred. No. 1.5e-38;
Matches 197; Conservative 86; Mismatches 215; Indels 50; Gaps 12;

QY 1 MRDYDEVTAFLEGGPQRFQ---LIFFLLSASIPNGFTGLSSVFLIATPEHR-CRYPDAAN 57
DB 1 MPTVDILQEVGESWFOKQAFILICLLLSAFAF---ICVGVFLGFTPDHHCQSPGVAE 57
QY 58 LSS-AWR-----NHTVPLRLRDGREVPHSCRRYR-----LATIA-NFSALG 96
DB 58 LSQRGWSPAEELNVTGICGPAEAFILQCRRRYEVDNQSALSCVDPLASLATNRSLP 117
QY 97 LEFGDVLQGLEQESCLDGEFSQDVYLSITVTENLVNLCEDDKWAKPLTISLFFVGVLLG 156
DB 118 LGP-----CQDGVY--DTPGSSIVTEFNLVCADSWKLDLFSQCLNAGFLFG 162
QY 157 SFISGQLDRFRGKRVLFVTVMQMTGFSFLQI-FSKNFEMFVFLVGMGQISNYVAAAFV 216
DB 163 SLGVGYFADRFGRKCLLGLTVLVNAVSGVLMAFSPNYKSMLLFRLLQLGLVSKGNMAGYT 222
QY 217 LGTEILGKSVRIIFSTLGVCIYAF--GYMVLPLFAFYFIRDRMMLLVALTMPGVLVALW 274
DB 223 LITEFVGSSR---RTVAIMYQMAFTVGLVLTGLAYALPHWRWLQAVSLPTFLFLYY 279
QY 275 WPIESPRWLISQGRFEAEVIRKAANGIVVPSTIFDPSSELDLSSKKQOQSHNILDLL 334
DB 280 WCVPESPRWLISQGRFEAEVIRKAANGIVVPSTIFDPSSELDLSSKKQOQSHNILDLL 337
QY 335 LRTNIRMTVMSIMLWMTISVGFGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLILL 394
DB 338 FTYPLRKRFTFLMTLWFTDVSIVYQGLIHMGATSGNLYLDLFLYSALVEICAFITALITI 397
QY 395 QYLPYRYSNATLALFGSVLLFMQLVPPDLYLATLVLMVCKFGVTAASVYVYTAELV 454
DB 398 DRVGRIYPMANSLLAGAACLVWIFISPDHLWLNIIICVGRMGITIAIQMCLVNAELY 457
QY 455 PTWVRNMGVSVSTASRLSGSILSPYFY--LGAYDRFLPYILMGSLTITAILTLFLPESF 513
DB 458 PTFVRNLGVWVCSLDCIGGIITPFIIVFRREVWQALPLILFAVLGLLAAGVTLLLPETK 517
QY 514 GTPLPDTI 521
DB 518 GVALPETM 525

RESULT 15
O15245
```

Search completed: January 25, 2002, 10:14:56  
Job time: 337 sec

```

ID O15245 PRELIMINARY; PRT; 554 AA.
AC O15245;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ORGANIC CATION TRANSPORTER.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA MEDLINE=97405886; PubMed=9260930;
RA Gorboulev V.G., Ulzheimer J.C., Akhondova A.;
RT "Cloning and characterization of two human polypeptide organic cation
RT transporters.";
RL DNA Cell Biol. 16:871-881(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL: X98332; CAA66977.1;
DR InterPro: IPR003662; sub transporter.
DR Pfam: PF00083; sugat_tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW Ion transport; Transmembrane.
SQ SEQUENCE 554 AA; 61187 MW; B16E200852CC4000 CRC64;

Query Match 25.58; Score 734; DB 4; Length 554;
Best Local Similarity 35.88; Pred. No. 2.7e-38;
Matches 196; Conservative 86; Mismatches 216; Indels 50; Gaps 12;

QY 1 MRDYDEVTAFLGEMGFQFQ---LTFELLSASIIPIPNFTGLSSVFLIATPEHRCRVPPDAAN 57
DB 1 MPTVDILLEQVGSFQKQAFLLICLLLSAFAF---ICVIVFLGTPDHCQSPGVAE 57

QY 58 LSS--AWR-----NHTVPLRLRDREVPHSCRRYR-----LATIA-NFSALG 96
DB 58 LSQRCGWSFAEELNTYVPLGLPAGEAFLGQCRRYEVDWNSQALSVDPLASLATNRSHLP 117

QY 97 LEPGRDVLGQLEQSCLDGNEFSQDVLSTIVTEWNLVCEDDKAPLTISLFFVGVILG 156
DB 118 LGP-----CQDGVY--DTPGSSIVTEFNLCADSWKLDLFQSLNAGFFFG 162

QY 157 SPISQSLDRFGKRVLFVTMGMTGFSFLQIFSKNFEMFVFLVYMGQISNYAAFY 216
DB 163 SLGVGYFADRFGRKICLLGLTVLNAVSGVLMAFSPNYMSMLLRLQLGLVSKGNMAGYT 222

QY 217 LGTEILGKSVRIIFSTLGVCFYAF--GYVVLPLFAFYFIRDRWMLLYALTMPGVLCVALW 274
DB 223 LITEFVGSGR---RTVAIMYQMAFTVGLVALTGLTAYALPHRWLQAVSLPTFLFLYY 279

QY 275 WPIESPRLWISQGFEEAEVIRKAAKANGIVVPSTIFDPSLQDLSSKKQOSHNIIDL 334
DB 280 WCVPEPRLWISQGFEEAEVIRKAAKANGIVVPSTIFDPSLQDLSSKKQOSHNIIDL 337

QY 335 LRTWNRMTTMSIMLMWTISVGYGLSLDTPNLHGDIYFNCFLSAMVEPAYVYLAWL 394
DB 338 FRTPLRKRRTFLMYLWFTDSVLYQGLLHMGATSGNLYLDLYSALVEIPGAFIALITI 397

QY 395 QYLPRYSNATALFGLGSVLLFMQLVPPDLYLATVLMVKGKFGVTAFAFSVYVYTAELY 454
DB 398 DRVGRIYPMAMNLLAGAACLVMIFFISPDHLWNLIIIMCVGRMGITIAIQMICLVNAELY 457

QY 455 PTVVRNMGVGSSTASRLGSLSPYFYV-LGAYDRELPYILMGSLTILTALTILFLPESF 513
DB 458 PTFVRLNMGVWVCSLCDIGGIITFEIVREWQALPLILFAVLGLAAGVTILLPETK 517

QY 514 GTPLPDTI 521
DB 518 GVALPETM 525

```



